



Repositorio Digital Institucional "José María Rosa"

Universidad Nacional de Lanús
Secretaría Académica
Dirección de Biblioteca y Servicios de Información Documental

Sebastian Martins

Derivación del proceso de explotación de información desde el modelado del negocio

Trabajo Final Integrador presentado para la obtención del título de Licenciado en Sistemas

Director de Trabajo Final Integrador

Ramón García Martínez

El presente documento integra el Repositorio Digital Institucional "José María Rosa" de la Biblioteca "Rodolfo Puiggrós" de la Universidad Nacional de Lanús (UNLa)

This document is part of the Institutional Digital Repository "José María Rosa" of the Library "Rodolfo Puiggrós" of the University National of Lanús (UNLa)

Cita sugerida

Martins, Sebastian. (2013). Derivación del proceso de explotación de información desde el modelado del negocio [en Línea]. Universidad Nacional de Lanús. Departamento de Desarrollo Productivo y Tecnológico

Disponible en: http://www.repositoriojmr.unla.edu.ar/descarga/TFI/LicSis/034626_Martins.pdf

Condiciones de uso

www.repositoriojmr.unla.edu.ar/condicionesdeuso



www.unla.edu.ar
www.repositoriojmr.unla.edu.ar
repositoriojmr@unla.edu.ar



DERIVACIÓN DEL PROCESO DE EXPLOTACIÓN DE INFORMACIÓN DESDE EL MODELADO DEL NEGOCIO

Alumno

APU Sebastian MARTINS

Directores

Dr. Ramón GARCIA-MARTINEZ y Mg Dario RODRIGUEZ

TRABAJO FINAL PRESENTADO PARA OBTENER EL GRADO
DE
LICENCIADO EN SISTEMAS

**DEPARTAMENTO
DE DESARROLLO PRODUCTIVO Y TECNOLÓGICO
UNIVERSIDAD NACIONAL DE LANUS**

DICIEMBRE, 2013

RESUMEN

Existen metodologías de explotación de información que destacan la importancia del planeamiento de una elicitación de requerimientos a lo largo de todo el proyecto de una manera ordenada, documentada, consistente y trazable. Sin embargo, dadas las características de este tipo de proyectos, el abordaje que propone la ingeniería de requerimientos clásica no es aplicable al proceso de identificar el problema de explotación de información, ni inferir el proceso que lo resuelve a partir del modelado del dominio del negocio. En este trabajo se propone un procedimiento basado en formalismos de Ingeniería del Conocimiento que permite derivar el proceso de explotación de información a partir de las representaciones del dominio de negocio y del problema de explotación de información.

ABSTRACT

There are information mining methodologies that emphasize the importance of planning for requirements elicitation along the entire project in an orderly, documented, consistent and traceable manner. However, given the characteristics of this type of project, the approach proposed by the classical requirements engineering is not applicable to the process of identifying the problem of information mining, nor allows to infer from the business domain modelling, the information mining process which solves it. In this grade thesis a method based on formalism of Knowledge Engineering is proposed, that allows deriving the information mining process from the representations of the business domain and the problem of information mining.

DEDICATORIA

A mi madre Silvia, a mi padre Juan, a mis abuelos Adela y Juan, y a mis hermanos Amilcar, Nayla y Melina por su apoyo incondicional, parte esencial de este logro académico

A mis amigos de la infancia Lucas, Rodrigo, Ignacio A., Ignacio D., Martin, Diego y Gastón, y a mi novia Daniela, por acompañarme durante todo este camino

A mi mentor Ramón, por su apoyo y confianza, y por ser un estímulo permanente en mi formación como persona y profesional.

AGRADECIMIENTOS

A la Universidad Nacional de Lanús por acogerme con generosidad de “alma máter” para que pudiera llevar a cabo mis estudios de Licenciatura en Sistemas

Al Grupo de Investigación en Sistemas de Información del Departamento de Desarrollo Productivo y Tecnológico de la Universidad Nacional de Lanús por recibirme cordialmente, proveyendo de un ambiente estimulante para mi formación

Al Dr. Ramón García-Martínez y al Mg Darío Rodríguez por dirigir mi trabajo final de licenciatura con la inquebrantable dedicación del maestro y el afecto del amigo

A mis profesores de carrera quienes me proveyeron de estímulos, respuestas y dudas para permitirme crecer profesional y personalmente

A mis colegas de carrera junto con los cuales crecí aprendiendo de los errores mutuos y del trabajo en equipo

A los colegas que cursaron Ingeniería del Software I en el primer cuatrimestre del presente año, quienes brindaron su apoyo para la consecución de este logro académico

A todos aquellos que generosamente me guiaron y me permitieron crecer en lo académico y en lo personal

ÍNDICE

1. INTRODUCCIÓN	1
1.1. MARCO DEL TRABAJO FINAL DE LICENCIATURA	1
1.2. DELIMITACIÓN DEL PROBLEMA	2
1.3. SOLUCIÓN PROPUESTA	2
1.4. VISIÓN GENERAL DEL TRABAJO FINAL DE LICENCIATURA	2
2. ESTADO DE LA CUESTIÓN	5
2.1. MARCO TEÓRICO DE LA EXPLOTACIÓN DE INFORMACIÓN DE INTERÉS PARA EL TRABAJO FINAL DE LICENCIATURA	5
2.1.1. Explotación de Información	5
2.1.2. Inteligencia de Negocios	6
2.1.3. Problemas de Negocio y de Explotación de Información	7
2.1.4. Procesos de Explotación de Información	8
2.1.4.1. Descubrimiento de Reglas de Comportamiento	8
2.1.4.2. Descubrimiento de Grupos	9
2.1.4.3. Ponderación de Interdependencia de Atributos	9
2.1.4.4. Descubrimiento de Reglas de Pertenencia a Grupos	11
2.1.4.5. Ponderación de Reglas de Comportamiento o de Pertenencia a Grupos	12
2.2. MARCO TEÓRICO DE INGENIERÍA DEL CONOCIMIENTO DE INTERÉS PARA EL TRABAJO FINAL DE LICENCIATURA	14
2.2.1. Redes Semánticas	14
2.2.2. Tabla Concepto-Categoría-Definición	15
2.2.3. Tabla Concepto-Atributo-Valor	15
3. DESCRIPCIÓN DEL PROBLEMA	17
3.1. IDENTIFICACIÓN DEL PROBLEMA DE INVESTIGACIÓN	17
3.2. PROBLEMA ABIERTO	19
3.3. SUMARIO DE INVESTIGACIÓN	19
4. SOLUCIÓN	21
4.1. PROCESO DE DERIVACIÓN DE MODELOS	21
4.1.1. Generalidades	21
4.1.2. Estructura General del Proceso de Derivación de Modelos	24
4.2. ESTANDARIZACIÓN DEL MODELADO	28
4.2.1. Notación General de Formalismos Basados en Redes Semánticas	29

4.2.2. Notación Específica de las Redes Semánticas del Problema de Explotación de Información	30
4.2.3. Resumen de Notación de Formalismos Basados en Redes Semánticas	31
4.3. PROCESO DE DERIVACIÓN DE MODELOS	31
4.3.1. Aspectos Generales	33
4.3.2. Fase Orientada al Análisis del Dominio del Negocio	34
4.3.2.1. Paso 1: “Identificación de los elementos y estructura del dominio”	35
4.3.2.1.1. Técnica Tabla Término-Categoría-Definición del Dominio	35
4.3.2.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio	37
4.3.2.2. Paso 2: “Identificación de Relaciones entre Conceptos del Dominio”	38
4.3.2.2.1. Técnica Tabla Concepto-Relación del Dominio	38
4.3.2.3. Paso 3: “Conceptualización Del Dominio”	39
4.3.2.3.1. Técnica Red Semántica del Modelo de Negocio	40
4.3.3. Fase Orientada al Análisis del Problema de Explotación de Información	41
4.3.3.1. Paso 4: “Identificación de los elementos y estructura del Problema de Explotación de Información”	41
4.3.3.1.1. Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información	42
4.3.3.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información	43
4.3.3.2. Paso 5: “Identificación de relaciones entre conceptos del Problema de Explotación de Información”	46
4.3.3.2.1. Técnica Tabla Concepto-Relación del Problema de Explotación de Información	46
4.3.3.3. Paso 6: “Conceptualización del Problema de Explotación de Información”	47
4.3.3.3.1. Técnica Red Semántica del Problema de Explotación de Información	48
4.3.4. Fase de Determinación del Proceso de Explotación de Información	49
4.3.4.1. Paso 7: “Derivación del Proceso de Explotación de Información”	49
4.3.4.1.1. Técnica Algoritmo de Derivación del Proceso de Explotación de Información	50
4.3.5. Resumen del Proceso de Derivación de Modelos	54
5. CASOS DE VALIDACIÓN	57
5.1. CASO DE VALIDACIÓN: EMPRESA DE PRESTACIÓN DE SERVICIOS DE INTERNET	57

5.1.1. Descripción	57
5.1.2. Aplicación de las Técnicas Utilizadas en la Fase de Análisis del Dominio del Negocio	58
5.1.2.1. Paso 1: “Identificación de los elementos y estructura del dominio”	58
5.1.2.1.1. Técnica Tabla Término-Categoría-Definición del Dominio	59
5.1.2.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio	60
5.1.2.2. Paso 2: “Identificación de Relaciones entre Conceptos del Dominio”	62
5.1.2.2.1. Técnica Tabla Concepto-Relación del Dominio	62
5.1.2.3. Paso 3: “Conceptualización Del Dominio”	63
5.1.2.3.1. Técnica Red Semántica del Modelo de Negocio	63
5.1.3. Aplicación de las Técnicas Utilizadas en la Fase de Análisis del Problema de Explotación de Información	64
5.1.3.1. Paso 4: “Identificación de los elementos y estructura del Problema de Explotación de Información”	64
5.1.3.1.1. Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información	65
5.1.3.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información	66
5.1.3.2. Paso 5: “Identificación de relaciones entre conceptos del Problema de Explotación de Información”	69
5.1.3.2.1. Técnica Tabla Concepto-Relación del Problema de Explotación de Información	69
5.1.3.3. Paso 6: “Conceptualización del Problema de Explotación de Información”	70
5.1.3.3.1. Técnica Red Semántica del Problema de Explotación de Información	70
5.1.4. Aplicación de la Técnica Utilizada en la Fase de Determinación del Proceso de Explotación de Información	72
5.1.4.1. Paso 7: “Derivación del Proceso de Explotación de Información”	72
5.1.4.1.1. Técnica Algoritmo de Derivación del Proceso de Explotación de Información	72
5.2. CASO DE VALIDACIÓN: INDUSTRIA AUTOMOTRIZ	74
5.2.1. Descripción	75
5.2.2. Aplicación de las Técnicas Utilizadas en la Fase de Análisis del Dominio del Negocio	75
5.2.2.1. Paso 1: “Identificación de los elementos y estructura del dominio”	76
5.2.2.1.1. Técnica Tabla Término-Categoría-Definición del Dominio	76

5.2.2.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio	78
5.2.2.2. Paso 2: “Identificación de Relaciones entre Conceptos del Dominio”	79
5.2.2.2.1. Técnica Tabla Concepto-Relación del Dominio	79
5.2.2.3. Paso 3: “Conceptualización Del Dominio”	80
5.2.2.3.1. Técnica Red Semántica del Modelo de Negocio	80
5.2.3. Aplicación de las Técnicas Utilizadas en la Fase de Análisis del Problema de Explotación de Información	82
5.2.3.1. Paso 4: “Identificación de los elementos y estructura del Problema de Explotación de Información”	82
5.2.3.1.1. Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información	82
5.2.3.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información	83
5.2.3.2. Paso 5: “Identificación de relaciones entre conceptos del Problema de Explotación de Información”	86
5.2.3.2.1. Técnica Tabla Concepto-Relación del Problema de Explotación de Información	86
5.2.3.3. Paso 6: “Conceptualización del Problema de Explotación de Información”	87
5.2.3.3.1. Técnica Red Semántica del Problema de Explotación de Información	87
5.2.4. Aplicación de la Técnica Utilizada en la Fase de Determinación del Proceso de Explotación de Información	90
5.2.4.1. Paso 7: “Derivación del Proceso de Explotación de Información”	90
5.2.4.1.1. Técnica Algoritmo de Derivación del Proceso de Explotación de Información	90
5.3. CASO DE VALIDACIÓN: MARKET BASKET ANALYSIS	92
5.3.1. Descripción	92
5.3.2. Aplicación de las Técnicas Utilizadas en la Fase de Análisis del Dominio del Negocio	93
5.3.2.1. Paso 1: “Identificación de los elementos y estructura del dominio”	93
5.3.2.1.1. Técnica Tabla Término-Categoría-Definición del Dominio	93
5.3.2.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio	94
5.3.2.2. Paso 2: “Identificación de Relaciones entre Conceptos del Dominio”	96
5.3.2.2.1. Técnica Tabla Concepto-Relación del Dominio	97
5.3.2.3. Paso 3: “Conceptualización Del Dominio”	98
5.3.2.3.1. Técnica Red Semántica del Modelo de Negocio	98

5.3.3. Aplicación de las Técnicas Utilizadas en la Fase de Análisis del Problema de Explotación de Información	100
5.3.3.1. Paso 4: “Identificación de los elementos y estructura del Problema de Explotación de Información”	100
5.3.3.1.1. Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información	100
5.3.3.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información	103
5.3.3.2. Paso 5: “Identificación de relaciones entre conceptos del Problema de Explotación de Información”	105
5.3.3.2.1. Técnica Tabla Concepto-Relación del Problema de Explotación de Información	106
5.3.3.3. Paso 6: “Conceptualización del Problema de Explotación de Información”	107
5.3.3.3.1. Técnica Red Semántica del Problema de Explotación de Información	107
5.3.4. Aplicación de la Técnica Utilizada en la Fase de Determinación del Proceso de Explotación de Información	109
5.3.4.1. Paso 7: “Derivación del Proceso de Explotación de Información”	109
5.3.4.1.1. Técnica Algoritmo de Derivación del Proceso de Explotación de Información	109
6. CONCLUSIONES	113
6.1. APORTES DEL TRABAJO FINAL DE LICENCIATURA	113
6.2. FUTURAS LÍNEAS DE INVESTIGACIÓN	115
7. REFERENCIAS	117

ÍNDICE DE FIGURAS

Figura 2.1	Proceso de explotación de información: descubrimiento de reglas de comportamiento [Britos y García-Martínez, 2009]	8
Figura 2.2	Proceso de explotación de información: descubrimiento de grupos [Britos y García-Martínez, 2009]	9
Figura 2.3	Proceso de explotación de información: ponderación de interdependencia de atributos [Britos y García-Martínez, 2009]	10
Figura 2.4	Proceso de explotación de información: descubrimiento de reglas de pertenencia a grupos [Britos y García-Martínez, 2009]	11
Figura 2.5	Proceso de explotación de información: ponderación de reglas de comportamiento o de pertenencia a grupos [Britos y García-Martínez, 2009]	13
Figura 2.6	Red Semántica	14
Figura 3.1	Representación gráfica de la brecha conceptual entre las actividades de determinación del negocio y del problema de explotación de información, y la definición del proceso de explotación de información a aplicar	19
Figura 4.1	Representación gráfica de la inserción del proceso de derivación de modelos propuesto entre las etapas de determinación el problema de negocio y del PEI, y del proceso de explotación de información	23
Figura 4.2	Estructura general del proceso de derivación de modelos, identificando sus tres fases orientadas al enfoque de las actividades, y sus entradas y salida	25
Figura 4.3	Flujo e interdependencia entre las fases, pasos y sus productos	26
Figura 4.4	Elemento para representar conceptos en redes semánticas	29
Figura 4.5	Elemento para representar atributos o valores en redes semánticas	29
Figura 4.6	Identificación de instancias de conceptos en redes semánticas	30
Figura 4.7	Representación de las distintas etiquetas que puede tener un arco según los elementos que relaciona. a) Relación entre un atributo y un valor, b) Relación de instanciación entre conceptos, c) Relación entre un concepto y un atributo y d) Relación de herencia entre conceptos	30
Figura 4.8	Elemento para representar atributos de entrada en redes semánticas	31
Figura 4.9	Elemento para representar atributos de salida en redes semánticas	31
Figura 4.10	Elemento para representar la variable identificadora del concepto como elemento de entrada en redes semánticas	31

Figura 4.11	Elemento para representar la variable identificadora del concepto como elemento de salida en redes semánticas	31
Figura 4.12	Red Semántica del Modelo de Negocio (RSMN) a partir del caso de estudio	41
Figura 4.13	Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de estudio	48
Figura 4.14	Identificación de subpaso 7.1 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de estudio	53
Figura 4.15	Identificación de subpaso 7.5 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de estudio	53
Figura 4.16	Identificación de subpaso 7.8 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de estudio	54
Figura 5.1	Red Semántica del Modelo de Negocio (RSMN) a partir del caso de validación: prestación de servicios de internet	64
Figura 5.2	Paso 1: Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: prestación de servicios de internet	71
Figura 5.3	Paso 2: Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: prestación de servicios de internet	71
Figura 5.4	Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: prestación de servicios de internet	72
Figura 5.5	Identificación de subpaso 7.1 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: prestación de servicios de internet	73
Figura 5.6	Identificación de subpaso 7.4 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: prestación de servicios de internet	74
Figura 5.7	Paso 1: Red Semántica del Modelo de Negocio (RSMN) a partir del caso de validación: industria automotriz	81
Figura 5.8	Red Semántica del Modelo de Negocio (RSMN) a partir del caso de validación: industria automotriz	81
Figura 5.9	Paso 1: Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: industria automotriz	88
Figura 5.10	Paso 2: Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: industria automotriz	89

Figura 5.11	Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: industria automotriz	89
Figura 5.12	Identificación de subpaso 7.1 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: industria automotriz	91
Figura 5.13	Identificación de subpaso 7.3 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: industria automotriz	91
Figura 5.14	Paso 1: Red Semántica del Modelo de Negocio (RSMN) a partir del caso de validación: market basket analysis	99
Figura 5.15	Red Semántica del Modelo de Negocio (RSMN) a partir del caso de validación: market basket analysis	99
Figura 5.16	Paso 1: Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: market basket analysis	108
Figura 5.17	Paso 2: Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: market basket analysis	108
Figura 5.18	Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: market basket analysis	109
Figura 5.19	Identificación de subpaso 7.1 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: market basket analysis	110
Figura 5.20	Identificación de subpaso 7.5 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: market basket analysis	111
Figura 5.21	Identificación de subpaso 7.6 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: market basket analysis	111

ÍNDICE DE TABLAS

Tabla 2.1	Tabla Concepto-Categoría-Definición	15
Tabla 2.2	Tabla Concepto-Atributo-Valor	16
Tabla 3.1	Tabla de carga de trabajo de cada fase del modelo	18
Tabla 4.1	Elementos propuestos y los modelos basados en redes semánticas a los que corresponden	32
Tabla 4.2	Código de regiones por provincias	34
Tabla 4.3	Técnica Tabla Término-Categoría-Definición del Dominio	36
Tabla 4.4	Tabla Término-Categoría-Definición del Dominio (TCDD) a partir del caso de estudio	36
Tabla 4.5	Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio	37
Tabla 4.6	Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) a partir del caso de estudio	38
Tabla 4.7	Técnica Tabla Concepto-Relación del Dominio	39
Tabla 4.8	Tabla Concepto-Relación del Dominio (CRD) a partir del caso de estudio	39
Tabla 4.9	Técnica Red Semántica del Modelo de Negocio	40
Tabla 4.10	Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información	43
Tabla 4.11	Tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) a partir del caso de estudio	44
Tabla 4.12	Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información	45
Tabla 4.13	Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) a partir del caso de estudio	45
Tabla 4.14	Técnica Tabla Concepto-Relación del Problema de Explotación de Información	47
Tabla 4.15	Tabla Concepto-Relación del Problema de Explotación de Información (CRPEI) a partir del caso de estudio	47
Tabla 4.16	Técnica Red Semántica del Problema de Explotación de Información	48
Tabla 4.17	Técnica Algoritmo de Derivación del Proceso de Explotación de Información	50

Tabla 4.18	Algoritmo de Derivación del Proceso de Explotación de Información (ADPEI)	51
Tabla 4.19	Resumen del Proceso de Derivación de Modelos	55
Tabla 5.1	Tabla Término-Categoría-Definición del Dominio (TCDD) a partir del caso de validación: prestación de servicios de internet	60
Tabla 5.2	Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) a partir del caso de validación: prestación de servicios de internet	62
Tabla 5.3	Tabla Concepto-Relación del Dominio (CRD) a partir del caso de validación: prestación de servicios de internet	63
Tabla 5.4	Tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) a partir del caso de validación: prestación de servicios de internet	66
Tabla 5.5	Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) a partir del caso de validación: prestación de servicios de internet	68
Tabla 5.6	Tabla Concepto-Relación del Problema de Explotación de Información (CRPEI) a partir del caso de validación: prestación de servicios de internet	70
Tabla 5.7	Tabla Término-Categoría-Definición del Dominio (TCDD) a partir del caso de validación: industria automotriz	77
Tabla 5.8	Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) a partir del caso de validación: industria automotriz	79
Tabla 5.9	Tabla Concepto-Relación del Dominio (CRD) a partir del caso de validación: industria automotriz	80
Tabla 5.10	Tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) a partir del caso de validación: industria automotriz	84
Tabla 5.11	Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) a partir del caso de validación: industria automotriz	86
Tabla 5.12	Tabla Concepto-Relación del Problema de Explotación de Información (CRPEI) a partir del caso de validación: industria automotriz	87
Tabla 5.13	Tabla Término-Categoría-Definición del Dominio (TCDD) a partir del caso de validación: market basket analysis	95
Tabla 5.14	Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) a partir del caso de validación: market basket analysis	97

Tabla 5.15	Tabla Concepto-Relación del Dominio (CRD) a partir del caso de validación: market basket analysis	98
Tabla 5.16	Tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) a partir del caso de validación: market basket analysis	102
Tabla 5.17	Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) a partir del caso de validación: market basket analysis	105
Tabla 5.18	Tabla Concepto-Relación del Problema de Explotación de Información (CRPEI) a partir del caso de validación: market basket analysis	107

NOMENCLATURA

ADPEI	Algoritmo de Derivación del Proceso de Explotación de Información
BI	Business Intelligence (Inteligencia de Negocios)
CARVD	Concepto-Atributo-Relación-Valor del Dominio
CARVD	Concepto-Atributo-Relación-Valor del Dominio
CARVEPEI	Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información
CAV	Concepto-Atributo-Valor
CCD	Concepto-Categoría-Definición
CRD	Concepto-Relación del Dominio
CRPEI	Concepto-Relación del Problema de Explotación de Información
DM	Data Mining (Minería de Datos)
GR	atributo clase de un grupo
MRI	Modelos de Representación Intermedia
OLAP	On Line Analytic Processing (Procesamiento Analítico en Línea)
PEI	Problema de Explotación de Información
RSMN	Red Semántica del Modelo de Negocio
RSPEI	Red Semántica del Problema de Explotación de Información
SOM	Self-Organizing Map (Mapas Auto Organizados)
TCDD	Término-Categoría-Definición del Dominio
TCDPEI	Término-Categoría-Definición del Problema de Explotación de Información
TDIDT	Top-Down Induction of Decision Trees (Inducción de Árboles de Decisión)
TFL	Trabajo Final de Licenciatura

1. INTRODUCCIÓN

En este capítulo se presenta el marco del Trabajo Final de Licenciatura (sección 1.1), se da una delimitación del problema (sección 1.2), se plantean los elementos de la solución propuesta (sección 1.3), y se da una visión general del proyecto (sección 1.4).

1.1. MARCO DEL TRABAJO FINAL DE LICENCIATURA

La inteligencia de negocio propone un abordaje interdisciplinario dentro del que se encuentra la Informática, que toma todos los recursos de información disponibles y el uso de herramientas analíticas y de síntesis con capacidad de transformar la información en conocimiento para generar, con base en estos recursos, conocimiento que contribuya con la toma de decisiones de gestión y generación de planes estratégicos en las organizaciones [Negash y Gray, 2008].

La explotación de información es la sub-disciplina de los sistemas de información que aporta a la inteligencia de negocio las herramientas para la transformación de información en conocimiento [García-Martínez et al., 2011]. Se define como la búsqueda de patrones interesantes y de regularidades importantes en grandes masas de información.

Un Proceso de Explotación de Información se define como un grupo de tareas relacionadas lógicamente [García-Martínez et al., 2013] que, a partir de un conjunto de información con un cierto grado de valor para la organización, se ejecuta para lograr otro, con un grado de valor mayor que el inicial. Adicionalmente, existe una variedad de técnicas de minería de datos, en su mayoría provenientes del campo del aprendizaje automático [García-Martínez et al., 2003], susceptibles de utilizar en cada uno de estos procesos.

En [Britos, 2008] se proponen los siguientes cinco procesos de explotación de información: (a) descubrimiento de reglas de comportamiento, que aplica cuando se requiere identificar cuáles son las condiciones para obtener determinado resultado en el dominio del problema; (b) descubrimiento de grupos, que aplica cuando se requiere identificar una partición en la masa de información disponible sobre el dominio de problema; (c) descubrimiento de atributos significativos, que aplica cuando se requiere identificar cuáles son los factores con mayor incidencia (o frecuencia de ocurrencia) sobre un determinado resultado del problema; (d) descubrimiento de reglas de pertenencia a grupos, que aplica cuando se requiere identificar cuáles son las condiciones de pertenencia a cada una de las clases en una partición desconocida “a priori”, pero presente en la masa de información disponible sobre el dominio de problema; (e) ponderación de reglas de

comportamiento o de pertenencia a grupos, que aplica cuando se requiere identificar cuáles son las condiciones con mayor incidencia (o frecuencia de ocurrencia) sobre la obtención de un determinado resultado en el dominio del problema, sean éstas las que, en mayor medida, inciden sobre un comportamiento o las que mejor definen la pertenencia a un grupo.

1.2. DELIMITACIÓN DEL PROBLEMA

Las metodologías de desarrollo de proyectos de explotación de información difieren de las usadas para los proyectos de desarrollo de software tradicionales. Las clásicas etapas de análisis, diseño, desarrollo, integración y pruebas no encajan con las etapas naturales de los procesos de desarrollo de este tipo de proyectos. En particular, las metodologías de explotación de información destacan la importancia de la comprensión del negocio y la vinculación entre el problema de negocio con el patrón o regularidad a descubrir en la información a partir del proceso de explotación de información a utilizar. Sin embargo, es un problema abierto determinar de manera sistemática cual es el proceso de explotación de información que se debe aplicar para resolver el problema de negocio.

1.3. SOLUCIÓN PROPUESTA

La solución propuesta tiene dos componentes: la identificación de formalismos de representación de conocimiento que permitan abordar el modelado del dominio de negocio y del problema de explotación de información, y un procedimiento que permita derivar el proceso de explotación de información a partir del modelado del dominio de negocio y del problema de explotación de información.

1.4. VISIÓN GENERAL DEL TFL

En el capítulo Introducción, se presenta el marco del Trabajo Final de Licenciatura, se da una delimitación del problema, se plantean los elementos de la solución propuesta, y se da una visión general del proyecto.

En el capítulo Estado de la Cuestión, se presentan los marcos teóricos de la explotación de información y de la ingeniería del conocimiento de interés para esta línea de investigación.

En el capítulo Descripción del Problema, se identifica el problema de investigación, en el cual se resalta la inexistencia de un mecanismo sistematizado para la determinación del proceso de

explotación de información a aplicar, se define el problema abierto y se concluye con un sumario de investigación.

En el capítulo Solución, se presenta un *proceso de derivación de modelos*, en el cual se abordan las cuestiones generales de mayor relevancia, se describe la estructura general del proceso propuesto. Se introducen los estándares definidos para el modelado de los distintos conocimientos, en las que se identifican las notaciones generales y específicas de los formalismos basados en redes semánticas y se presenta un resumen de los mismos. Por último, se presenta una visión detallada del proceso de derivación de modelos propuesto, en la cual se describen los aspectos generales del proceso y el caso de estudio mediante el cual se ilustrará al mismo, se definen las fases que componen al proceso, se identifican las etapas que componen cada una de las fases, se señalan las técnicas y formalismos que se utilizan para el desarrollo de cada fase y se brinda un resumen del proceso.

En el Capítulo Casos de Validación, se introducen tres casos de validación los cuales presentan distintas características a partir de los cuales se aplican las fases, etapas y técnicas que componen al *proceso de derivación de modelos*. Se analiza un primer caso de validación correspondiente a una empresa de prestación de servicios de internet, la cual busca mejorar la permanencia de sus clientes. se analiza un segundo caso de validación correspondiente a un negocio perteneciente a la industria automotriz, cuyo objetivo principal es la detección e identificación de daños y/o averías producidas a lo largo del circuito que recorren los autos desde que salen de la planta de producción hasta que llegan a su destino final. Se analiza el tercer caso de validación correspondiente a un análisis de la canasta de compras (“market basket analysis”), en el cual se desea identificar las dependencias entre los productos.

En el Capítulo Conclusiones, se presentan las aportaciones de este Trabajo Final de Licenciatura y se destacan las futuras líneas de investigación que se consideran de interés en base al problema abierto que se presenta en este trabajo.

En el Capítulo Referencias se listan todas las publicaciones consultadas para el desarrollo de esta investigación.

2. ESTADO DE LA CUESTIÓN

En este capítulo se presenta el estado de la cuestión sobre distintas teorías y técnicas que convergen con los objetivos del presente trabajo. Se presenta el marco teórico de la explotación de información de interés para el Trabajo Final de Licenciatura (sección 2.1), el cual se describen los conceptos: explotación de información (sección 2.1.1), inteligencia de negocios (sección 2.1.2), problema de negocio y de explotación de información (sección 2.1.3) y procesos de explotación de información (sección 2.1.4), en la que se cada uno de los cinco procesos: descubrimiento de reglas de comportamiento (Sección 2.1.4.1), descubrimiento de grupos (Sección 2.1.4.2), descubrimiento de atributos significativos (Sección 2.1.4.3), descubrimiento de reglas de pertenencia a grupos (Sección 2.1.4.4) y ponderación de reglas de comportamiento o de pertenencia (Sección 2.1.4.5); y el marco teórico de la ingeniería del conocimiento de interés para esta línea de investigación (sección 2.2), en el cual se identifican los siguientes conceptos: redes semánticas (sección 2.2.1), tabla Concepto-Categoría-Definición (sección 2.2.2) y tabla Concepto-Atributo-Valor (sección 2.2.3).

2.1. MARCO TEÓRICO DE LA EXPLOTACIÓN DE INFORMACIÓN DE INTERÉS PARA EL TFL

El marco teórico de la explotación de información de interés para el trabajo final de la carrera, se estructura en base al marco teórico que se presenta en [Britos, 2008], identificando los siguientes conceptos: explotación de información (sección 2.1.1), inteligencia de negocios (sección 2.1.2), problema de negocio y de explotación de información (sección 2.1.3) y procesos de explotación de información (sección 2.1.4).

2.1.1. Explotación de Información

La explotación de información se ha definido, aunque utilizando el termino Minería de datos en su lugar, como la búsqueda de patrones interesantes y de reglas importantes en grandes masas de información [Fayad et al., 1996; Grossman et al., 1999]. También fue definida como el proceso de descubrir nuevas correlaciones, patrones y tendencias significativas utilizando grandes cantidades de datos almacenados en repositorios, usando tecnologías de reconocimiento de patrones, así como técnicas matemáticas y de estadística [Larose, 2005].

La Ingeniería de Explotación de Información estudia los procesos de extracción de conocimiento no trivial que recibe de manera implícita en los datos que se almacenan en grandes bases de datos. Estos procesos se valen de algoritmos de Minería de Datos (DM, Data Mining) [García-Martínez et al., 2011]. Dicho conocimiento es previamente desconocido y puede resultar útil para algún proceso [Stefanovic et al., 2006]. Para un experto, o para el responsable de un sistema, normalmente no son los datos en sí lo más relevante, sino el conocimiento oculto presente en sus relaciones, fluctuaciones y dependencias.

2.1.2. Inteligencia de Negocios

Se denomina inteligencia de negocio (Business Intelligence, BI) al conjunto de estrategias y herramientas [Reinschmidt y Allison, 2000] enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización [Koubarakis y Plexousakis, 2000; Grigori et al., 2001; Grigori, et al., 2004; Mäkipää, 2004; Negash, 2008].

Involucra el uso de los datos de una organización para facilitarle a las personas que realizan la toma de decisiones estratégicas del negocio, la comprensión del funcionamiento actual y la anticipación de acciones para dar una dirección bien informada a la organización [Thomas, 2001; 2003; Lönnqvist y Pirttimäki, 2006].

El sistema de información que se utiliza para la inteligencia de negocios integra diferentes datos extraídos del área de producción, con información relacionada con la organización o sus ámbitos y con datos económicos [Schiefer et al., 2004].

Las herramientas de inteligencia de negocios son software de aplicación diseñados para colaborar con la inteligencia de negocios en los procesos de las organizaciones [Pirttimäki et al., 2006].

Específicamente se trata de herramientas que asisten el análisis y la presentación de los datos [Reiter et al., 2000; Morik y Rüping, 2002; Moss, 2003; Rudin y Cressy, 2003; Nguyen et al., 2005].

Este conjunto de herramientas y metodologías tienen en común las siguientes características [Osterwalder, 2004]:

- **Accesibilidad a la información:** Los datos son la fuente principal de este concepto. Lo primero que deben garantizar este tipo de herramientas y técnicas será el acceso de los usuarios a los datos, con independencia de la procedencia de éstos.
- **Apoyo en la toma de decisiones:** Se busca ir más allá en la presentación de la información, de manera que los usuarios tengan acceso a herramientas de análisis que les permitan seleccionar y manipular sólo aquellos datos que les interesen.

- **Orientación al usuario final:** Se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas.

De acuerdo a su nivel de complejidad, las herramientas de inteligencia de negocios se pueden clasificar en [Langseth y Vivatrat, 2003; Dasgupta y Vankayala, 2007]:

- **Cuadro de Mando Integral:** También conocidos como Business Intelligence Dashboards, o Dashboards Ejecutivos. Son resúmenes visuales de información de la organización, que pretende mostrar de una mirada la comprensión global de las condiciones de la organización mediante métricas e indicadores clave de desempeño (KPIs). Esta es una herramienta de inteligencia de negocios en uso desde hace unos pocos años.
- **Procesamiento Analítico en Línea (OLAP / On Line Analytic Processing):** Es la capacidad de algunos sistemas de soporte de decisiones gerenciales, que permiten examinar de manera interactiva grandes volúmenes de información desde varias perspectivas.
- **Aplicaciones de Reportes:** Genera vistas de datos agregadas para mantener a la gestión informada sobre el estado de la organización.
- **Explotación de Información:** Extracción de información de las bases de datos acerca del objeto observado, mediante la utilización de aplicaciones que pueden aislar e identificar patrones o tendencias del objeto observado en un alto volumen de datos. Hay una gran variedad de técnicas de Explotación de Información que revelan distintos tipos de patrones.

2.1.3. Problemas de Negocio y de Explotación de Información

Para lograr el éxito de un proyecto de explotación de información es necesario comprender lo que el cliente desea lograr, para ello se deben entender los problemas de negocio que se presentan.

Los problemas de negocio describen los principales objetivos del cliente desde una perspectiva del negocio, además de otras cuestiones relacionadas con el objetivo principal del negocio que el cliente desea abordar [Chapman, et al., 1999].

Los problemas de negocio manifiestan los objetivos en la terminología del negocio. Los problemas de explotación de información representan los objetivos del proyecto en términos técnicos. La respuesta esperada obtenida del problema de explotación de información permite lograr los objetivos del negocio [Chapman, et al., 1999].

2.1.4. Procesos de Explotación de Información

Los procesos de explotación de información definen las técnicas o algoritmos a utilizar en base a las características del problema de explotación. En [Britos, 2008] se presentan los siguientes procesos de explotación de información: descubrimiento de reglas de comportamiento (Sección 2.1.4.1), descubrimiento de grupos (Sección 2.1.4.2), descubrimiento de atributos significativos (Sección 2.1.4.3), descubrimiento de reglas de pertenencia a grupos (Sección 2.1.4.4) y ponderación de reglas de comportamiento o de pertenencia (Sección 2.1.4.5).

2.1.4.1. Descubrimiento de Reglas de Comportamiento

En [Britos y García-Martínez, 2009] se define que el proceso de descubrimiento de reglas de comportamiento aplica cuando se requiere identificar cuáles son las condiciones para obtener determinado resultado en el dominio del problema. Son ejemplos de problemas que requieren este proceso: identificación de características del local más visitado por los clientes, identificación de factores que inciden en el alza las ventas de un producto dado, establecimiento de características o rasgos de los clientes con alto grado de fidelidad a la marca, establecimiento de atributos demográficos y psicográficos que distinguen a los visitantes de un website, entre otros. Para el descubrimiento de reglas de comportamiento definidos a partir de atributos clases en un dominio de problema que representa la masa de información disponible, se propone la utilización de algoritmos de inducción TDIDT para descubrir las reglas de comportamiento de cada atributo clase. Este proceso y sus subproductos pueden ser visualizados gráficamente en la figura 2.1. Como resultado de la aplicación del algoritmo de inducción TDIDT al atributo clase se obtiene un conjunto de reglas que definen el comportamiento de dicha clase.

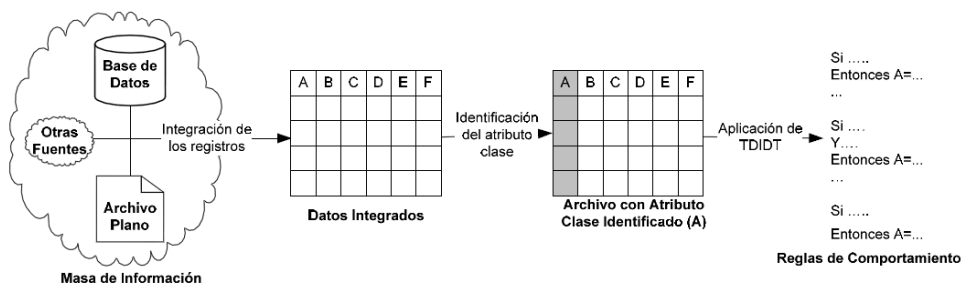


Figura 2.1. Proceso de explotación de información: descubrimiento de reglas de comportamiento [Britos y García-Martínez, 2009]

2.1.4.2. Descubrimiento de Grupos

En [Britos y García-Martínez, 2009] se define que el proceso de descubrimiento de grupos aplica cuando se requiere identificar una partición en la masa de información disponible sobre el dominio de problema. Son ejemplos de problemas que requieren este proceso: identificación de segmentos de clientes para bancos y financieras, identificación de tipos de llamadas de clientes para empresas de telecomunicación, identificación de grupos sociales con las mismas características, identificación de grupos de estudiantes con características homogéneas, entre otros. Para el descubrimiento de grupos a partir de masas de información del dominio de problema sobre las que no se dispone ningún criterio de agrupamiento “a priori” se propone la utilización de Mapas Auto Organizados de Kohonen o SOM por su sigla en inglés. El uso de esta tecnología busca descubrir si existen grupos que permitan una partición representativa del dominio de problema que la masa de información disponible representa. Este proceso y sus subproductos pueden ser visualizados gráficamente en la figura 2.2.

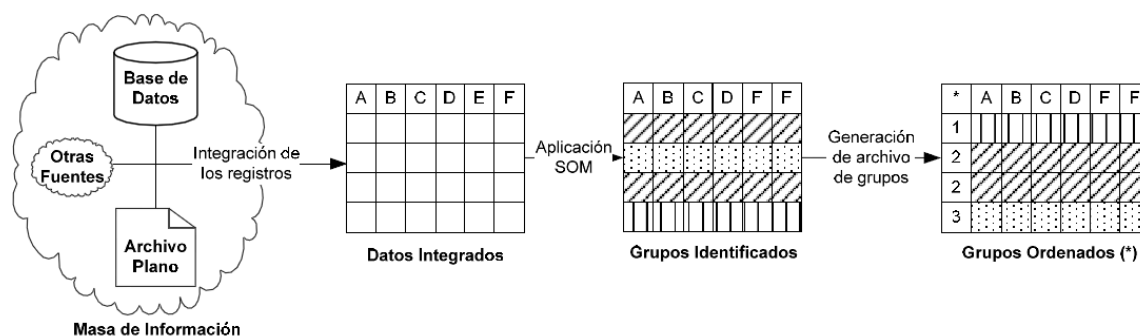


Figura 2.2. Proceso de explotación de información: descubrimiento de grupos [Britos y García-Martínez, 2009]

En primer lugar se identifican todas las fuentes de información (bases de datos, archivos planos, entre otras), se integran entre sí formando una sola fuente de información a la que se llamará datos integrados. Con base en los datos integrados se aplican mapas auto-organizados (SOM). Como resultado de la aplicación de SOM se obtiene una partición del conjunto de registros en distintos grupos a los que se llamará grupos identificados. Para cada grupo identificado se generará el archivo correspondiente.

2.1.4.3. Ponderación de Interdependencia de Atributos

En [Britos y García-Martínez, 2009] se define que el proceso de ponderación de interdependencia de atributos aplica cuando se requiere identificar cuáles son los factores con mayor incidencia (o frecuencia de ocurrencia) sobre un determinado resultado del problema. Son ejemplos de problemas

que requieren este proceso: factores con incidencia sobre las ventas, rasgos distintivos de clientes con alto grado de fidelidad a la marca, atributos claves que convierten en vendible a un determinado producto, características sobresalientes que tienen los visitantes de un website, entre otros. Para ponderar en que medida la variación de los valores de un atributo incide sobre la variación del valor de un atributo clase se propone la utilización de Redes Bayesianas. El uso de esta tecnología busca identificar si existe interdependencia en algún grado entre los atributos que modelan el dominio de problema que la masa de información disponible representa. Este proceso y sus subproductos pueden ser visualizados gráficamente en la figura 2.3. En primer lugar se identifican todas las fuentes de información (bases de datos, archivos planos, entre otras), se integran entre sí formando una sola fuente de información a la que se llamará datos integrados. Con base en los datos integrados se selecciona el atributo clase (atributo A en la Figura).

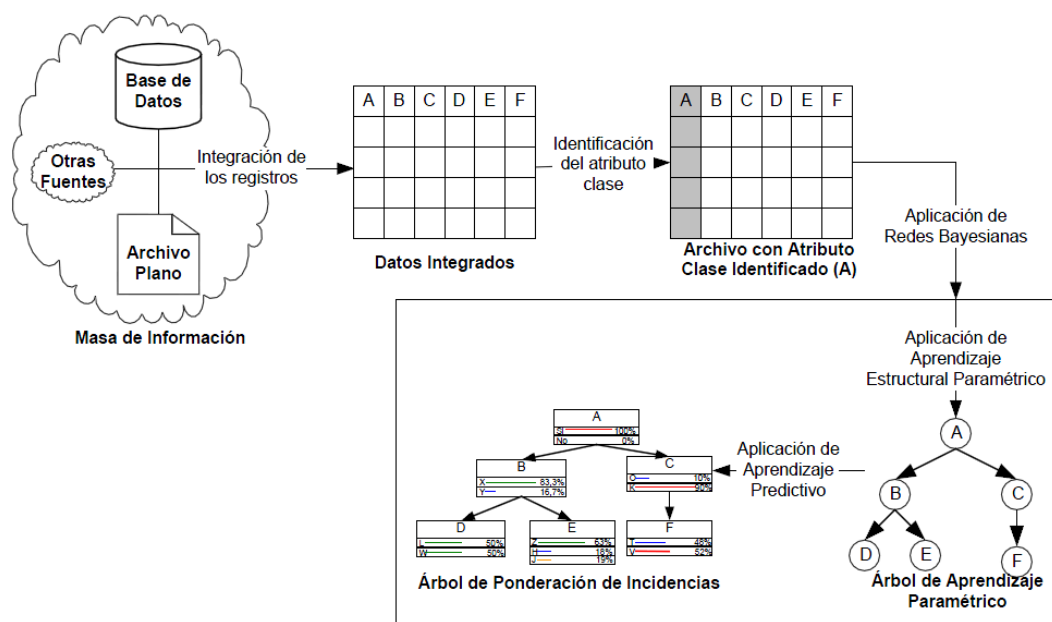


Figura 2.3. Proceso de explotación de información: ponderación de interdependencia de atributos [Britos y García-Martínez, 2009]

Como resultado de la aplicación del aprendizaje estructural de las Redes Bayesianas al archivo con atributo clase identificado se obtiene el árbol de aprendizaje; a este se le aplica el aprendizaje predictivo Redes Bayesianas y se obtiene el árbol de ponderación de interdependencias que tiene como raíz al atributo clase y como nodos hojas a los otros atributos con la frecuencia (incidencia) sobre el atributo clase.

2.1.4.4. Descubrimiento de Reglas de Pertenencia a Grupos

En [Britos y García-Martínez, 2009] se define que el proceso de descubrimiento de reglas de pertenencia a grupos aplica cuando se requiere identificar cuáles son las condiciones de pertenencia a cada una de las clases en una partición desconocida “a priori”, pero presente en la masa de información disponible sobre el dominio de problema. Son ejemplos de problemas que requieren este proceso: tipología de perfiles de clientes y caracterización de cada tipología, distribución y estructura de los datos de mi website, segmentación etaria de mis estudiantes y comportamiento de cada segmento, clases de llamadas telefónicas en una región y caracterización de cada clase, entre otros. Para el descubrimiento de reglas de pertenencia a grupos se propone la utilización de mapas auto-organizados (SOM) para el hallazgo de los mismos y; una vez identificados los grupos, la utilización de algoritmos de inducción (TDIDT) para establecer las reglas de pertenencia a cada uno. Este proceso y sus subproductos pueden ser visualizados gráficamente en la figura 2.4.

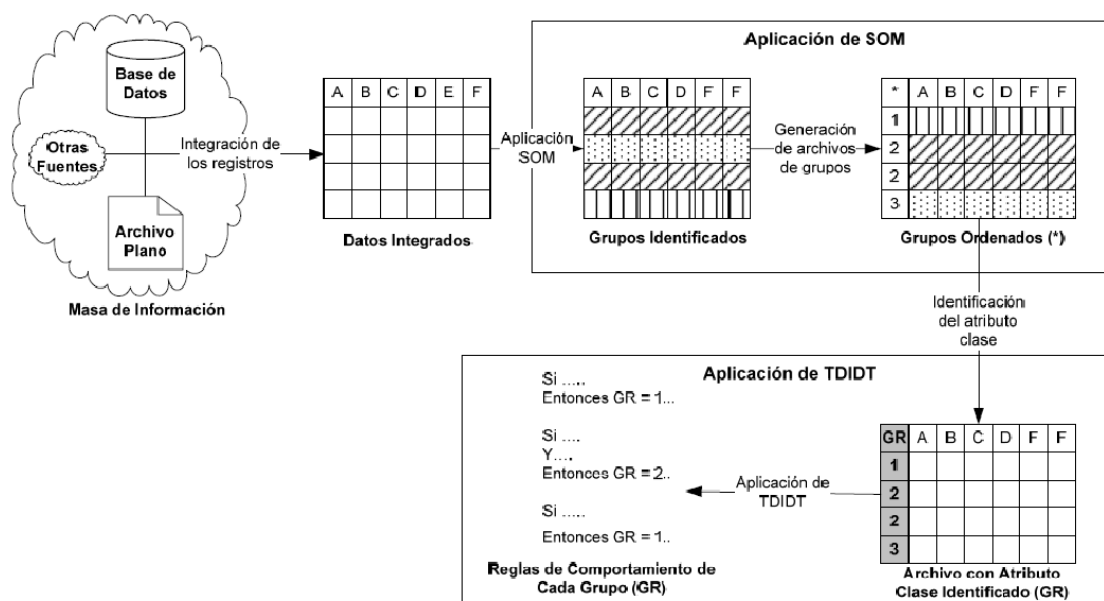


Figura 2.4. Proceso de explotación de información: descubrimiento de reglas de pertenencia a grupos [Britos y García-Martínez, 2009]

En primer lugar se identifican todas las fuentes de información (bases de datos, archivos planos, entre otras), se integran entre sí formando una sola fuente de información a la que se llamará datos integrados. Con base en los datos integrados se aplican mapas auto-organizados (SOM). Como resultado de la aplicación de SOM se obtiene una partición del conjunto de registros en distintos grupos a los que se llama grupos identificados. Se generan los archivos asociados a cada grupo identificado. A este conjunto de archivos se lo llama grupos ordenados. El atributo “grupo” de cada

grupo ordenado se identifica como el atributo clase de dicho grupo, constituyéndose este en un archivo con atributo clase identificado (GR). Se aplica el algoritmo de inducción TDIDT al atributo clase de cada grupo GR y se obtiene un conjunto de reglas que definen el comportamiento de cada grupo.

2.1.4.5. Ponderación de Reglas de Comportamiento o de Pertenencia a Grupos

En [Britos y García-Martínez, 2009] se define que el proceso de ponderación de reglas de comportamiento o de la pertenencia a grupos aplica cuando se requiere identificar cuáles son las condiciones con mayor incidencia (o frecuencia de ocurrencia) sobre la obtención de un determinado resultado en el dominio del problema, sean estas las que en mayor medida inciden sobre un comportamiento o las que mejor definen la pertenencia a un grupo. Son ejemplos de problemas que requieren este proceso: identificación del factor dominante que incide en el alza las ventas de un producto dado, rasgo con mayor presencia en los clientes con alto grado de fidelidad a la marca, frecuencia de ocurrencia de cada perfil de clientes, identificación del tipo de llamada más frecuente en una región, entre otros.

Para la ponderación de reglas de comportamiento o de pertenencia a grupos se propone la utilización de redes bayesianas. Esto puede hacerse a partir de dos procedimientos dependiendo de las características del problema a resolver: cuando no hay clases/grupos identificados; o cuando hay clases/grupos identificados. El procedimiento a aplicar cuando hay clases/grupos identificados consiste en la utilización de algoritmos de inducción TDIDT para descubrir las reglas de comportamiento de cada atributo clase y posteriormente se utiliza redes bayesianas para descubrir cuál de los atributos establecidos como antecedentes de las reglas tiene mayor incidencia sobre el atributo establecido como consecuente. Este proceso y sus subproductos pueden ser visualizados gráficamente en la figura 2.5.

En primer lugar se identifican todas las fuentes de información (bases de datos, archivos planos, entre otras), se integran entre sí formando una sola fuente de información a la que se llamará datos integrados. Con base en los datos integrados se selecciona el atributo clase (atributo A en la figura 2.5). Como resultado de la aplicación del algoritmo de inducción TDIDT al atributo clase se obtiene un conjunto de reglas que definen el comportamiento de dicha clase. Seguidamente, se construye un archivo con los atributos antecedentes y consecuentes identificados por la aplicación del algoritmo TDIDT.

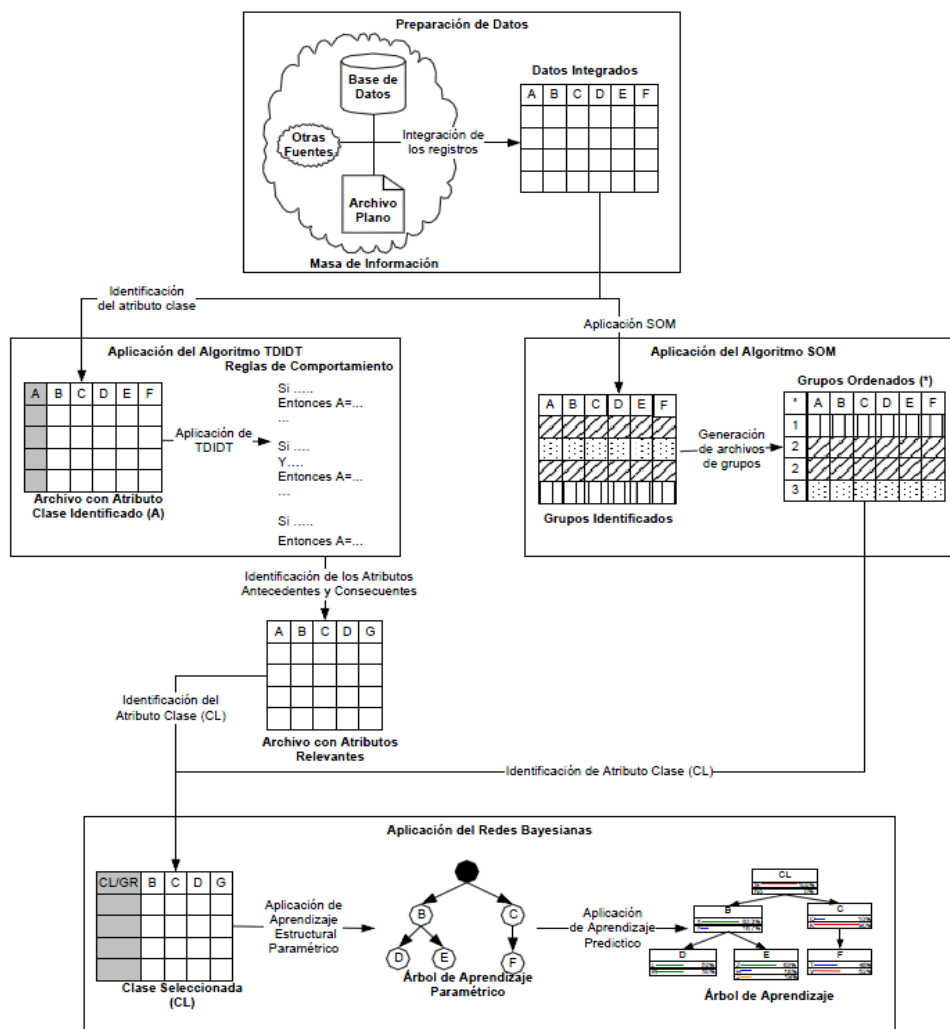


Figura 2.5. Proceso de explotación de información: ponderación de reglas de comportamiento o de pertenencia a grupos [Britos y García-Martínez, 2009]

Como resultado de la aplicación del aprendizaje estructural de las Redes Bayesianas al archivo con atributo clase obtenido por la utilización del algoritmo TDIDT (en la figura 2.5), se obtiene el árbol de aprendizaje; a este se le aplica aprendizaje predictivo y se obtiene el árbol de ponderación de interdependencias que tiene como raíz al atributo clase (en este caso el atributo consecuente) y como nodos hojas a los atributos antecedentes con la frecuencia (incidencia) sobre el atributo consecuente. El procedimiento a aplicar cuando no hay clases/grupos identificados consiste en identificar todas las fuentes de información (bases de datos, archivos planos, entre otras), se integran entre sí formando una sola fuente de información a la que se llamará datos integrados. Con base en los datos integrados se aplican mapas auto-organizados (SOM). Como resultado de la aplicación de SOM se obtiene una partición del conjunto de registros en distintos grupos a los que se llamará grupos identificados. Para cada grupo identificado se generará el archivo correspondiente. A este conjunto de archivos se lo llama grupos ordenados. El atributo “grupo” de

cada grupo ordenado se identifica como el atributo clase de dicho grupo, constituyéndose este en un archivo con atributo clase identificado (GR). Como resultado de la aplicación del aprendizaje estructural se obtiene el árbol de aprendizaje; a este se le aplica el aprendizaje predictivo y se obtiene el árbol de ponderación de interdependencias que tiene como raíz al atributo grupo y como nodos hojas a los otros atributos con la frecuencia (incidencia) sobre el atributo grupo.

2.2. MARCO TEÓRICO DE INGENIERÍA DEL CONOCIMIENTO DE INTERÉS PARA EL TFL

Dentro del marco teórico de ingeniería del conocimiento de interés para el trabajo final de la carrera, se identifican los siguientes conceptos: redes semánticas (sección 2.2.1), tabla Concepto-Categoría-Definición (sección 2.2.2) y tabla Concepto-Atributo-Valor (sección 2.2.3).

2.2.1. Redes Semánticas

En [Sowa, 1992] se define a las redes semánticas como una notación grafica para representar conocimiento a través de patrones de nodos y arcos interconectados. Asimismo se las define como un formalismo o paradigma de representación de conocimiento basado en relaciones, las cuales constan de un grafo orientado formado por nodos etiquetados que representan conceptos o entidades y arcos unidireccionales etiquetados que representan relaciones entre los conceptos o instancias [Winston, 1994]. El formalismo de captura es un grafo orientado como se exhibe en la figura 2.6.

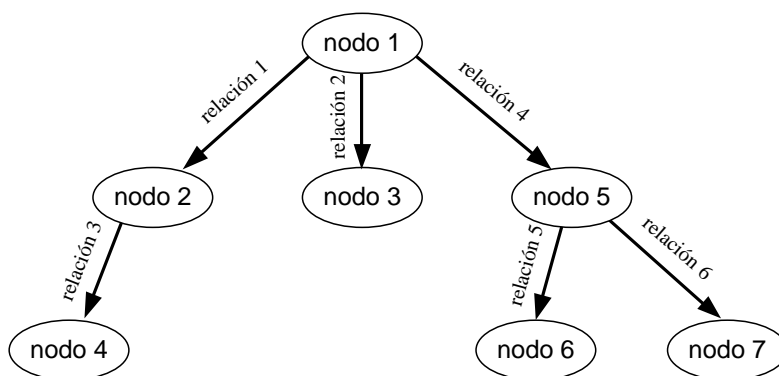


Figura 2.6. Red Semántica

2.2.2. Tabla Concepto-Categoría-Definición

En [Rodríguez y García Martínez, 2012] se presenta un conjunto de formalismos para modelar las iteraciones en espacios virtuales para trabajos colaborativos. Para este trabajo surge el interés en la tabla Concepto-Categoría-Definición.

La función de la tabla Concepto-Categoría-Definición (CCD) es representar los conocimientos fácticos del modelo conceptual de dinámica grupal. En la tabla CCD se introducen en orden lexicográfico los conceptos que se utilizaran en los otros dos formalismos especificando la categoría y dando la definición. Un concepto puede ser de alguna de las siguientes categorías: actor, objeto y acción. Los actores son los sujetos de la dinámica grupal. Los objetos son la materia o asunto que recibe el ejercicio de las facultades de accionar de los actores. Las acciones definen procesos que los actores ejecutan sobre objetos o sobre otros actores. El formalismo de captura es el de una tabla como se muestra en la tabla 2.1.

Concepto	Categoría	Definición
Concepto 1	Categoría 1	Definición del Concepto 1
Concepto 2	Categoría 1	Definición del Concepto 2
-----	-----	-----
Concepto N	Categoría Q	Definición del Concepto N

Tabla 2.1. Tabla Concepto-Categoría-Definición

2.2.3. Tabla Concepto-Atributo-Valor

En el marco de los modelos para representar conocimientos que propone [García Martínez y Britos, 2004] se presenta la tabla Concepto-Atributo-Valor (CAV) como herramienta para capturar conocimientos.

La tabla Concepto-Atributo-Valor (CAV), es una técnica que modela conocimientos de tipo fácticos, esta proporciona una lista de los conceptos que se manipulan en el dominio de conocimiento relacionado con la familia de problemas que resolverá el Sistema Experto a desarrollar. Cada concepto quedará descrito en términos de los atributos que definen a cada concepto y de los valores que cada atributo puede tomar. La representación genérica de dicho formalismo se ilustra en la tabla 2.2.

Concepto	Atributo	Valor
Concepto 1	Atributo 1	Valor 1, valor 2, ... , valor N
	Atributo 2	Valor 1, valor 2, ... , valor N

	Atributo N	Valor 1, valor 2, ... , valor N
...	Atributo 1	Valor 1, valor 2, ... , valor N
	Atributo 2	Valor 1, valor 2, ... , valor N
Concepto N	Atributo 1	Valor 1, valor 2, ... , valor N
	Atributo 2	Valor 1, valor 2, ... , valor N

Tabla 2.2. Tabla Concepto-Atributo-Valor

3. DESCRIPCIÓN DEL PROBLEMA

En este capítulo se presenta el problema de investigación a analizar en el trabajo final de licenciatura en sistemas, comenzando con la identificación del problema de investigación (sección 3.1), posteriormente se caracteriza el problema abierto (sección 3.2) y se concluye con un sumario de investigación (sección 3.3).

3.1. IDENTIFICACIÓN DEL PROBLEMA DE INVESTIGACIÓN

Se ha definido una metodología para la gestión y desarrollo de sistemas de explotación de información (Modelo de Procesos para Proyectos de Explotación de Información [Vanrell, 2011]), en dicha metodología se establecen etapas destinadas al entendimiento del negocio, así como a determinar los objetivos del proceso de explotación de información que dichos problemas conllevan.

Dichos procesos son esenciales para el éxito del proyecto, omitir el esfuerzo necesario para la comprensión detallada del problema de negocio conllevará a obtener la respuesta correcta a un problema equivocado o que no satisfaga las demandas del cliente, mientras que evitar una correcta determinación de las metas de la explotación de información producirá una respuesta incorrecta al problema en cuestión.

Además de los riesgos recién mencionados, es sustancial destacar que las etapas de entendimiento del negocio y de modelado son las etapas que mayor tiempo requieren en un proyecto de explotación de información bajo el proceso de desarrollo del proyecto del modelo previamente mencionado.

En [Rodríguez et al., 2010], se describen los tiempos requeridos por cada fase del proyecto, y detalla las fases y subfases que hacen un mayor uso de dicho recurso. A partir de los resultados obtenidos, se destaca que entre las fases de *Comprensión del Negocio* y de *Modelado* se invierte más del 50% de la duración del proyecto. Además, las subfases *Determinar los Objetivos de Negocio* y *Evaluar la Situación*, que integran la fase de *Comprensión del Negocio*, utilizan más del 70% del tiempo pautado. Por otro lado, en la fase de *Modelado*, la subfase *Construcción del Modelo*, requiere el 62,97% del plazo concedido a la totalidad de la ejecución de la fase a la que pertenece. En la tabla 3.1 se presenta el porcentaje de tiempo que cada fase conlleva:

FASE	% del TIEMPO
Fase 1 COMPRENSIÓN DEL NEGOCIO	20,70
Fase 2 ENTENDIMIENTO DE LOS DATOS	10,90
Fase 3 PREPARACION DE DATOS	15,61
Fase 4 MODELADO	34,41
Fase 5 EVALUACIÓN	7,45
Fase 6 IMPLANTACION	10,93

Tabla 3.1. *Tabla de carga de trabajo de cada fase del modelo*

A partir de comprender lo crítico de dichas etapas, se resalta que ninguna de las metodologías para proyectos de explotación de información existentes ofrece un método que permita de forma estandarizada definir el proceso de explotación de información a aplicar, en base al dominio del negocio y al problema de explotación de información identificado, debiendo cada ingeniero del conocimiento, representar de la forma que coincidirá oportuna dichos conocimientos, dificultando la comprensión de los mismos.

Mediante la creación de un proceso, y la estandarización en la representación del negocio y del problema de explotación de información, se favorece en la documentación y la comprensión de los proyectos, permitiendo identificar similitudes entre distintos proyectos, reduciendo los tiempos de los mismos y favoreciendo la mejora de los distintos procesos productivos pertenecientes al ciclo de vida del producto.

A partir de la necesidad de utilizar un formalismo para modelar los conocimientos previamente mencionados, surge la necesidad de establecer un conjunto de reglas que estandaricen el modelado de los conocimientos involucrados en el proceso.

Los problemas previamente mencionados pueden ser sintetizados de la siguiente manera:

- Se carece de una forma estandarizada de representar los problemas de negocio y de explotación de información.
- Se carece de una terminología y de una semántica definida para el modelado del dominio del negocio y problema de explotación de información.
- Se carece de un procedimiento que permita establecer el proceso de explotación de información a utilizar, a partir del dominio del negocio, y del problema de explotación de información.

3.2. PROBLEMA ABIERTO

En la presente sección se identifica el problema abierto, que surge de la necesidad de determinar una forma estandarizada de llevar a cabo las actividades pertenecientes a las etapas críticas de entendimiento del negocio y de modelado, con el fin de garantizar la calidad y el éxito del proyecto. Además, la ejecución de estas actividades de manera desestructurada y sin estándares, obstaculiza la mejora de las mismas en proyectos futuros. A su vez, surge la necesidad de definir un conjunto de normas para la modelización de los conocimientos del dominio de negocio y del problema de explotación de información.

Por consiguiente y en virtud de todo lo expuesto, el problema abierto que se aborda en este trabajo final de carrera, está caracterizado por la brecha conceptual existente entre la definición del modelo de negocio y del problema de explotación de información, y la determinación del proceso de explotación de información a aplicar para dicho contexto (representada gráficamente en la figura 3.1). Además se destaca la carencia de una forma estándar de representar los conocimientos pertenecientes al modelo de negocio y al problema de explotación de información.

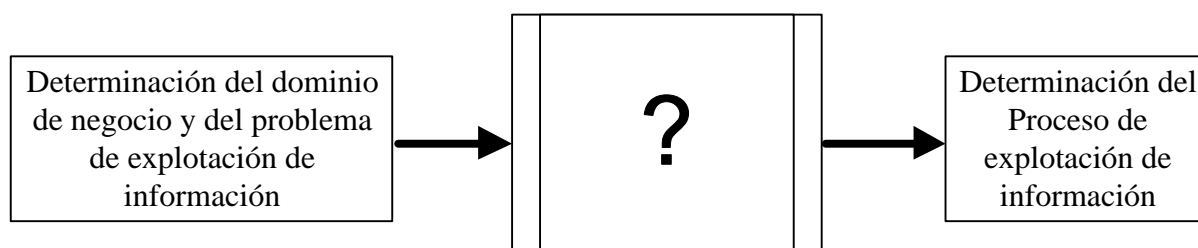


Figura 3.1. Representación gráfica de la brecha conceptual entre las actividades de determinación del dominio del negocio y del problema de explotación de información, y la definición del proceso de explotación de información a aplicar

3.3. SUMARIO DE INVESTIGACIÓN

De lo anteriormente expuesto surgen las siguientes preguntas de investigación:

Pregunta 1: ¿Existen formalismos que permitan representar el conocimiento del negocio y del problema de explotación de información, mediante los cuales se puedan estandarizar y derivar conclusiones sobre el proceso de explotación de información que provee la solución al problema planteado? En caso afirmativo: ¿Cuáles?

Pregunta 2: ¿Pueden identificarse entre distintos modelos de negocio y problemas de información, patrones que permitan identificar el proceso de explotación de información que satisface al problema analizado? En caso afirmativo: ¿Cuáles?

Pregunta 3: En caso de existir patrones que identifiquen el proceso de explotación de información que satisface al problema analizado, ¿puede definirse un proceso que estandarice la derivación del proceso de explotación de información a partir de los conocimientos previamente mencionados? En caso de existir: ¿Cuáles son sus actividades?

4. SOLUCIÓN

En el presente capítulo se presenta un modelo de proceso de derivación de modelos (sección 4.1), del cual se abordan las cuestiones generales de mayor relevancia (sección 4.1.1), se presenta la estructura general de la propuesta de dicho modelo (sección 4.1.2). Se presentan los estándares definidos para el modelado (sección 4.2), dentro de esta se distingue la Notación General de Formalismos Basados en Redes Semánticas (sección 4.2.1) y Notación Específica de Formalismos Basados en Redes Semánticas (sección 4.2.2), y se expone un resumen de las notaciones propuestas para formalismos basados en redes semánticas (sección 4.2.3). Se presenta una visión detallada del *proceso de derivación de modelos* (sección 4.3), en la cual se abarcan los aspectos generales del *proceso de derivación de modelos* (sección 4.3.1), se describe cada una de las fases que componen dicho proceso y sus etapas: análisis del dominio del negocio (sección 4.3.2), análisis del problema de explotación de información (sección 4.3.3) y determinación del proceso de explotación de información (sección 4.3.4), y se presenta un resumen del mismo (sección 4.3.5).

4.1. PROCESO DE DERIVACIÓN DE MODELOS

En esta sección se presenta una propuesta de modelo de *proceso de derivación de modelos*, estructurada en dos partes: generalidades (sección 4.1.1) y la estructura general de la propuesta del modelo de *proceso de derivación de modelos* (sección 4.1.2).

4.1.1. Generalidades

En función del análisis realizado en el capítulo 3 correspondiente a la descripción del problema, se considera de interés citar nuevamente el problema abierto que se aborda en este trabajo final de la carrera, el cual se focaliza en la brecha conceptual existente en la transición del proceso de explotación de información a aplicar, a partir de la descripción del dominio de negocio, y del problema de explotación de información. Esta brecha conceptual, dificulta la ejecución de la primeras cuatro etapas del Proceso de Desarrollo del Proyecto de Explotación de Información (de acuerdo a la modelo propuesto por [Vanrell, 2011]), teniendo un impacto crítico en el resultado final del proyecto.

A su vez, la definición de un proceso lógico para la transformación de las características del negocio y del problema de explotación de información al proceso de explotación de información, brinda una serie de actividades bien estructuradas y resultados claramente definidos, que favorecen en la comprensión de las distintas actividades que el proyecto conlleva, así como un mayor

entendimiento de la estructura de los datos relevantes para el mismo, favoreciendo el desarrollo de las etapas de entendimiento y preparación de los datos.

En las siguientes citas podemos comprobar la importancia de estandarizar en la ingeniería de software:

“La estandarización de procesos provocará que de manera coordinada los procesos y esfuerzos se diseñen de forma común, es decir, todos los departamentos de la empresa o bien si se ella cuenta con otras localidades o centros de trabajo entenderán y verán los mismo, este lenguaje único permitirá mejorar la comunicación y dará soporte en todo momento a la toma de decisiones. La organización que pretenda una evolución y crecimiento controlado podrá tomar esta iniciativa.”

[Sánchez, 2012]. Del párrafo de Sánchez, se desprende que un proceso estándar mejora la comunicación entre los miembros del proyecto, reduciendo el esfuerzo y el tiempo necesario para el desarrollo de las distintas actividades, además de producirse una documentación homogénea a través de los distintos proyectos que la empresa u organización lleve a cabo, facilitando su uso.

Los procesos del software se pueden mejorar por la estandarización del proceso donde la diversidad de los procesos del software en una organización sea reducida. Esto conduce a mejorar la comunicación y a una reducción del tiempo de formación, y hace la ayuda al proceso automatizado más económica. La estandarización también es un primer paso importante para introducir nuevos métodos, técnicas y buenas prácticas de ingeniería del software. Esta mejora en los procesos significa entender los procesos existentes y mejorar la calidad del producto y/o reducir los costos y tiempos de desarrollo [Sommerville, 2005].

A partir de lo recién mencionado, se propone como solución insertar una actividad entre las etapas de determinación del modelo de negocio, y la etapa de determinación del problema de explotación de información y la definición del proceso de explotación de información, la cual denominaremos *Proceso de Derivación de Modelos* (figura 4.1), y cuya función es, actuar como enlace entre las etapas determinación del negocio y determinación del problema de explotación de información (PEI), y la etapa de definición del proceso de explotación de información, eliminando la arbitrariedad en las actividades que ellas involucran, favoreciendo el éxito de las mismas.

El objetivo del *proceso de derivación de modelos* es proveer de un mecanismo estandarizado para representar la descripción del dominio de negocio y la descripción del problema de explotación de información, en modelos, a partir de los cuales se determina cuál de los procesos de explotación de información satisface o brinda la solución al problema definido, favoreciendo de este modo, al éxito y calidad de los resultados del proyecto, y dejando constancia del proceso implementado a través de la documentación obtenida, proporcionando información relevante para proyectos posteriores.

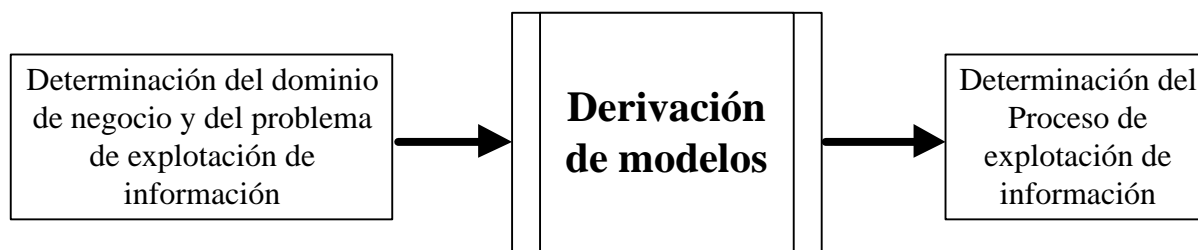


Figura 4.1. Representación gráfica de la inserción del proceso de derivación de modelos propuesto entre las etapas de determinación del dominio de negocio y del PEI, y del proceso de explotación de información

A partir de la implementación de esta actividad de derivación de modelos es posible la consecución de un conjunto de representaciones gráficas denominadas *Modelos de Representación Intermedia (MRI)* a partir de las cuales es posible representar la información contenida en la descripción del dominio de negocio y la descripción del problema de explotación de información (por lo general en formato de “lenguaje natural” y es así como se la supone presentada en este trabajo), a los efectos de simplificar la comprensión del proceso de explotación de información a aplicar, mediante la identificación de patrones o estructuras similares detectables en las representaciones realizadas.

Estos modelos de representación intermedia, hacen uso de distintos formalismos de Ingeniería de Conocimiento, entre ellos: las Redes Semánticas [Sowa, 1992], las tablas Concepto-Atributo-Valor [García Martínez y Britos, 2004] y la tabla Concepto-Categoría-Definición propuesta en [Rodríguez, 2012].

Estas representaciones intermedias estarán conformadas, fundamentalmente, por un conjunto de representaciones gráficas basadas en las redes semánticas: Red Semántica del Modelo de Negocio y Red Semántica del Problema de Explotación de Información.

Para la consecución del objetivo principal del trabajo, mediante la aplicación de las representaciones gráficas basadas en las redes semánticas, surge la necesidad de definir estándares para la representación de dichos formalismos, definiendo un marco de trabajo común, obteniendo así todos los beneficios que esto implica.

En síntesis, este trabajo final de carrera propone el *proceso de derivación de modelos*, en el cual se define su estructura de forma general (sección 4.1.2) y su estructura de forma detallada (sección 4.3), como herramienta para la derivación de la descripción del dominio de negocio y del problema de explotación de información, en el proceso de explotación de información a aplicar, y propone una serie de reglas aplicadas al modelado de la información del negocio, con el objetivo de estandarizar los modelos de representación intermedia y la documentación que estos producen, explicado en detalle en la sección 4.2.

4.1.2. Estructura General del Proceso de Derivación de Modelos

La actividad de derivación de modelos se lleva a cabo por medio de un proceso denominado *proceso de derivación de modelos*, el cual se estructura en tres fases, de acuerdo al enfoque de las actividades que las componen:

1. Una primera fase *orientada al análisis del dominio del negocio*, cuyo objetivo es el análisis, la comprensión y la conceptualización del modelo de negocio de la organización. Dicha fase se realizará por única vez para los distintos problemas de negocio que la organización desea enfrentar.
2. Una segunda fase *orientada al análisis del problema de explotación de información*, cuyo objetivo abarca el análisis, la comprensión y el modelado del problema de explotación de información. Dicha fase deberá realizarse por cada problema de negocio que se quiera afrontar.
3. Una tercera fase *orientada a determinar el proceso de explotación de información*, cuyo objetivo es identificar el proceso de explotación de información a aplicar, de acuerdo al dominio del negocio y al problema de explotación de información, expresados mediante las representaciones intermedias obtenidas en las fases previas.

Este proceso toma como elementos de partida la descripción del dominio del negocio, dentro de la cual se encuentran definidos los datos del negocio, y la descripción del problema de explotación de información, junto con la descripción de sus datos, y proporciona como salida el proceso de explotación de información a aplicar para dar solución al problema de negocio representado mediante el problema de explotación de información. La figura 4.2 ilustra este concepto.

El soporte principal del *proceso de derivación de modelos* está compuesto por tres fases, cuyas primeras dos están conformadas por tres etapas, y la última por una.

Las tres etapas que conforman la primera fase, están orientadas al modelado del dominio del negocio y cuyo resultado final es la Red Semántica del Modelo de Negocio, que brinda una representación de los elementos que componen el dominio del negocio y la interacción entre los mismos.

La segunda fase orientada al modelado del problema de explotación de información, está compuesta por tres actividades cuyo resultado final es la Red Semántica del Problema de Explotación de Información, que brinda una representación de los elementos que componen el dominio del negocio y la interacción entre los mismos.

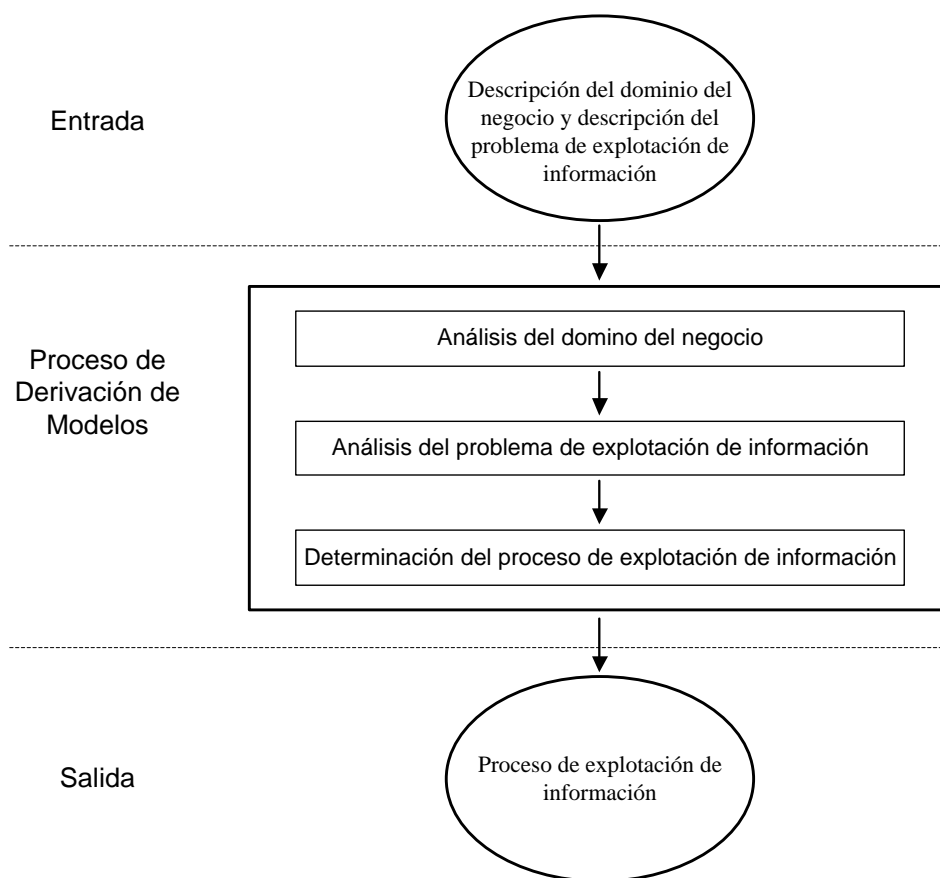


Figura 4.2. Estructura general del proceso de derivación de modelos, identificando sus tres fases orientadas al enfoque de las actividades, y sus entradas y salida

La tercera fase orientada a determinar el proceso de explotación de información, está conformada por una actividad la cual se compone de ocho subpasos, y su salida es el proceso de explotación de información a aplicar, es decir, el resultado final esperado por el *proceso de derivación de modelos*. La distribución de las fases no es de forma arbitraria, sino que definen el orden en el que se deben realizar cada una de las actividades que las conforman. Cada paso o etapa del proceso, tiene elementos de entrada, los cuales se procesan, generando elementos de salida, siendo estos a su vez, elementos de entrada de actividades posteriores.

En la figura 4.3 se ilustra el flujo de pasos que conlleva la ejecución del *proceso de derivación de modelos*, las interdependencias entre las fases, sus etapas y sus productos, representadas mediante el ingreso como elemento de entrada de un paso, uno o más elementos de salida de otro. En dicha figura se presentan las siete etapas del proceso de derivación de modelos, distribuidas según las fases a las que pertenecen. Dichas etapas se conectan mediante flechas las cuales indican los elementos de entrada (aquellas que convergen con la parte izquierda del paso) y salida (aquellas que surgen del sector derecho del paso) correspondientes a cada etapa y como cada elemento de salida retroalimenta otra etapa posterior.

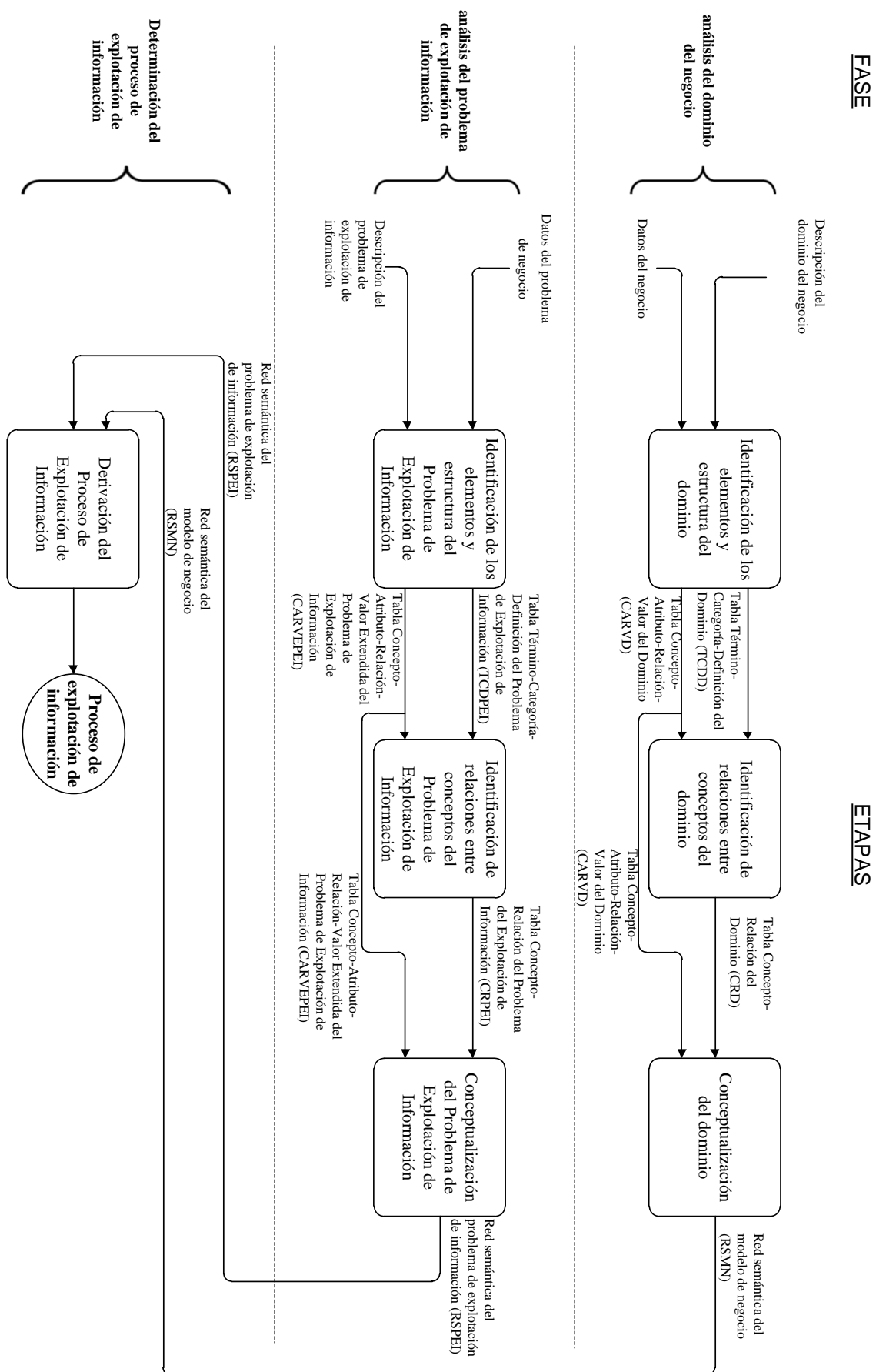


Figura 4.3. Flujo e interdependencia entre las fases, pasos y sus productos

El orden de ejecución de los pasos se define a continuación, identificando las etapas con negrita, los nombre de los pasos entre comillas y los elementos de entrada/salida con cursiva:

El primer paso a ejecutar es el perteneciente a la fase orientada al **análisis del dominio del negocio**, denominado “Identificación de los elementos y estructura del dominio”, el cual tiene como elementos de entrada la *descripción del dominio del negocio* y los *datos del negocio*, y produce como salida la *tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD)* y la *tabla Término-Categoría-Definición del Dominio (TCDD)*, los cuales son elementos de entrada al segundo paso denominado “Identificación de relaciones entre conceptos del dominio”, y cuyo elemento resultado es la *tabla Concepto-Relación del Dominio (CRD)*, la cual es utilizada por el tercer y último paso de la fase “Conceptualización del dominio”, junto con la *tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD)* como elementos de entrada, produciendo como resultado final de la etapa (pero no del proceso), la *Red Semántica del Modelo de Negocio (RSMN)*, la cual será utilizada posteriormente.

Dentro de la fase orientada al **análisis del problema de explotación de información**, podemos observar que los tres pasos que la componen, son independientes a la fase anteriormente descrita, por lo cual su ejecución puede ser realizada en paralelo.

La primera etapa que debe ejecutarse es “Identificación de los elementos y estructura del Problema de Explotación de Información”, la cual tiene como elementos de entrada los *datos del problema de negocio* y la *descripción del problema de explotación de información*, produciendo como resultado de aplicar dicho paso, dos modelos de representación intermedia, la *tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI)* y la *tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI)*.

Ambos elementos mencionados como resultados del paso anterior, ingresan como elementos de entrada a la etapa de “Identificación de relaciones entre conceptos del Problema de Explotación de Información”, la cual proporciona como producto de salida la *tabla Concepto-Relación del Problema de Explotación de Información (CRPEI)*, que junto con la *tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI)*, alimentan al último paso de la etapa, denominado “Conceptualización del Problema de Explotación de Información”, y que proporciona como resultado la *Red Semántica del Problema de Explotación de Información (RSPEI)*, que será posteriormente utilizada.

A partir de la conceptualización del dominio del negocio y del problema de explotación de información mediante las redes semánticas, obtenidas a partir de la ejecución del tercer paso correspondiente a las dos fases previas respectivamente, se poseen los elementos necesarios para comenzar con la última fase del proceso. Esta fase está compuesta por un único paso denominado “Derivación del Proceso de Explotación de Información”, cuyos elementos de entrada son la *Red*

semántica del modelo de negocio (RSMN) y la Red semántica del problema de explotación de información (RSPEI), y a partir de los cuales, aplicando una serie de reglas se obtiene como resultado el Proceso de explotación de información que satisface el problema de explotación de información analizado.

4.2. ESTANDARIZACIÓN DEL MODELADO

En esta sección, se definen un conjunto de reglas a aplicar para la realización de los modelos que surgen como producto de las distintas etapas del proceso propuesto, destinados a representar el dominio de negocio y los problemas de explotación de información.

El objetivo de definir una serie de reglas para la representación de dichos problemas es el de establecer un estándar o marco de trabajo, que permita mantener la trazabilidad entre distintos proyectos, facilitando la comprensión de los mismos.

El formalismo definido para la representación de los modelos son las redes semánticas, definidas en el capítulo 2.

Este formalismo se centra en las relaciones existentes entre los conceptos, siendo ampliamente utilizados para representar bases y/o modelos de conocimientos. Entre sus ventajas se puede destacar la simplicidad para graficar el conocimiento, lo que también se ve reflejado en una rápida e intuitiva comprensión del conocimiento que este representa.

La simplicidad para modelar el conocimiento está dada por el reducido número de símbolos y la facilidad visual para interpretar el conocimiento representado.

A pesar de las ventajas descritas, las redes semánticas carecen de una terminología universalmente aceptada y de una semántica uniforme y precisa, dicha afirmación puede verse manifestado para el formalismo de las redes semánticas en el siguiente párrafo perteneciente a [Suárez-Figueroa y Gómez-Pérez, 2008]:

“Se puede utilizar cualquier etiqueta para dar nombre a cualquier nodo o arco. Los principales problemas de este formalismo de representación son la falta de una terminología adecuada y universalmente aceptada y de una semántica uniforme y precisa”.

Para dar solución a dicho problema, se propone un conjunto de reglas para estandarizar el formalismo aplicado al modelado del negocio y a los problemas de explotación de información, con el objetivo de establecer un marco de trabajo común, para su aplicación en el *proceso de derivación de modelos*.

De acuerdo a lo presentado en la sección anterior, se distinguen aquellos elementos que hacen uso de las reglas propuestas, dichos modelos son:

- la Red Semántica del Modelo de Negocio, y

- la Red Semántica del Problema de Explotación de Información.

Entre el conjunto de reglas definidas, podemos segmentarlas en dos grupos, el primero que define los elementos de representación comunes en los distintos modelos basados en redes semánticas (sección 4.2.1), y el segundo orientado a aquellos elementos particulares de la Red Semántica del Problema de Explotación de Información, que identifican los datos de entrada y salida del problema (sección 4.2.2). En la sección 4.2.3 se presenta un resumen de los elementos propuestos y los modelos a los que aplican.

4.2.1. Notación General de Formalismos Basados en Redes Semánticas

A continuación se definen las reglas de representación que aplican a los ambos modelos indicados en la sección 4.2:

1. Los conceptos se representan mediante óvalos con sus líneas continuas (Figura 4.4),
2. los atributos o valores se representan a través de óvalos con sus líneas intermitentes (Figura 4.5),
3. el identificador de una instancia del concepto se define dentro del ovalo concatenándolo al tipo de concepto con un guion “-” entre ambos (Figura 4.6),
4. el valor o referencia de la etiqueta de los arcos dependerá de la relación que esta represente:
 - a. Si une un atributo con su valor, se etiqueta con la palabra “valor” (Figura 4.7.a),
 - b. Si indica una relación de instancia de un concepto, se etiqueta con la palabra “instancia de” (Figura 4.7.b),
 - c. Si actúa de enlace entre un concepto y su atributo, se etiqueta con una o más palabras que representen la conexión entre ambos (Figura 4.7.c),
 - d. Si indica una relación de herencia de un concepto, se etiqueta con la palabra “subclase de” (Figura 4.7.d).



Figura 4.4. Elemento para representar conceptos en redes semánticas

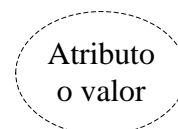


Figura 4.5. Elemento para representar atributos o valores en redes semánticas

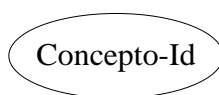


Figura 4.6. Identificación de instancias de conceptos en redes semánticas

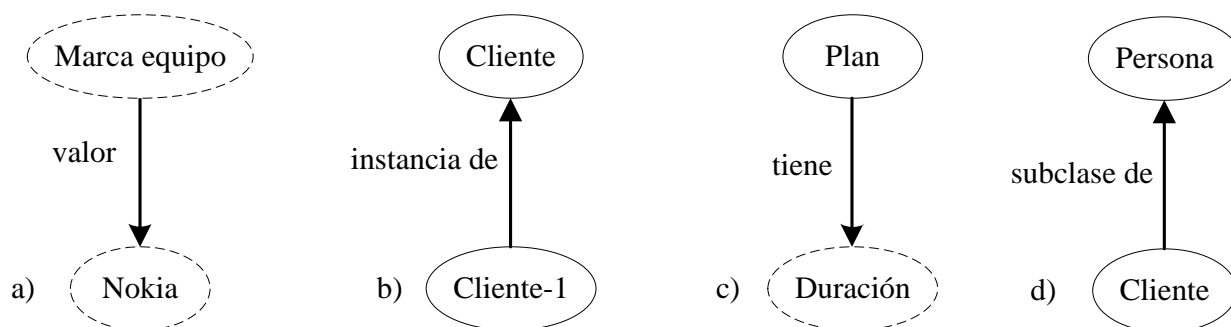


Figura 4.7. Representación de las distintas etiquetas que puede tener un arco según los elementos que relaciona. **a)** Relación entre un atributo y un valor, **b)** Relación de instanciación entre conceptos, **c)** Relación entre un concepto y un atributo y **d)** Relación de herencia entre conceptos

Es primordial destacar de la última figura presentada (Figura 4.7), que en los casos “a)”, “b)” y “d)”, el valor posible de la etiqueta se encuentra definido, sin aceptarse otro valor para dichas relaciones, mientras que en el caso “c)”, se propone un valor ejemplo, ya que dependiendo del contenido de los conceptos-atributos que se relacionen, el valor de la etiqueta del arco variará.

La primera imagen de la figura 4.7, indica que hay un atributo “Marca equipo”, cuyo valor es Nokia. La segunda imagen, indica que el concepto “cliente-1” (siendo 1 el identificador del cliente como se explicó en la figura 4.6) pertenece al concepto general cliente por lo tanto obtiene todas las características que a este lo representa. La tercera imagen, indica una relación de pertenencia entre el concepto plan y el atributo duración, indicando que todo plan tiene una duración estimada. La última imagen representa una relación de herencia, en el que la clase hereda todas las características que definen al concepto padre, en este caso el concepto Cliente hereda todas sus características de Persona, ya que un cliente es una persona, de acuerdo a esta relación.

4.2.2. Notación Específica de las Redes Semánticas del Problema de Explotación de Información

En la Red Semántica del Problema de Explotación de Información, se debe identificar el flujo del problema descrito, a través de la identificación de los elementos de entrada y de salida, para ello se incorporan elementos adicionales, para poder indicar dicha información. Dichas notaciones adicionales, son:

1. Los atributos de entrada se representan con doble ovalo con sus líneas intermitentes (Figura 4.8),
2. Los atributos de salida se representan con doble ovalo, teniendo el ovalo interno su línea continua y el ovalo externo su línea intermitente (Figura 4.9),
3. Los identificadores de instancias de conceptos, que sean elemento de entrada se representan mediante doble ovalo, teniendo el ovalo externo su línea continua, y el ovalo interno su línea intermitente (Figura 4.10).
4. Los identificadores de instancias de conceptos, que sean elemento de salida se representan mediante doble ovalo cuyas líneas son continuas (Figura 4.11).

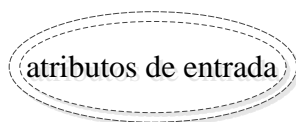


Figura 4.8. Elemento para representar atributos de entrada en redes semánticas

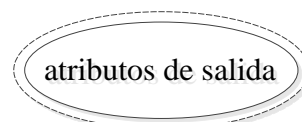


Figura 4.9. Elemento para representar atributos de salida en redes semánticas

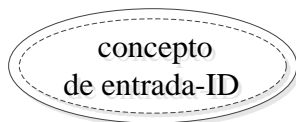


Figura 4.10. Elemento para representar la variable identificadora del concepto como elemento de entrada en redes semánticas

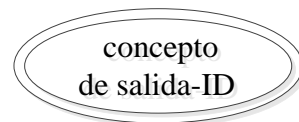


Figura 4.11. Elemento para representar la variable identificadora del concepto como elemento de salida en redes semánticas

4.2.3. Resumen de Notación de Formalismos Basados en Redes Semánticas

La tabla 4.1 que sintetiza los estándares definidos para los modelos basados en redes semánticas, y a que modelos corresponden.

4.3. PROCESO DE DERIVACIÓN DE MODELOS

En esta sección, se describe en detalle al *proceso de derivación de modelos* propuesto, brindando una visión ampliada de la provista en la sección 4.1, y haciendo uso de los formalismos definidos en la sección 4.2.


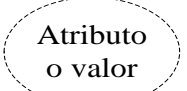


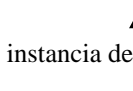

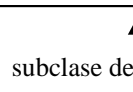
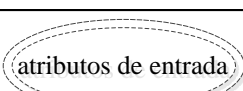
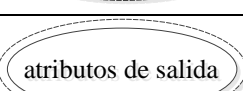
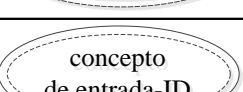
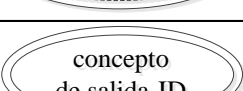
Elemento notacional	Formalismo donde es utilizado	
	Red Semántica del Modelo de Negocio (RSMN)	Red Semántica del problema de explotación de Información (RSPEI)
	✓	✓
	✓	✓
	✓	✓
	✓	✓
	✓	✓
	✓	✓
	✓	✓
		✓
		✓
		✓
		✓

Tabla 4.1. Elementos propuestos y los modelos basados en redes semánticas a los que corresponden

Se presentan los aspectos generales del proceso propuesto y se describe el caso de estudio con el que se ilustrará el mismo (sección 4.3.1), se describe detalladamente la fase *orientada al análisis del dominio del negocio* (sección 4.3.2), y se definen los pasos que componen dicha fase: “Identificación de los elementos y estructura del dominio” (sección 4.3.2.1), “Identificación de relaciones entre conceptos del dominio” (sección 4.3.2.2) y “Conceptualización del dominio” (sección 4.3.2.3). Se explica de forma semejante la fase *orientada al análisis del problema de explotación de información* (sección 4.3.3), la que se describe a partir de las etapas que la

componen: “Identificación de los elementos y estructura del Problema de Explotación de Información” (sección 4.3.3.1), “Identificación de relaciones entre conceptos del Problema de Explotación de Información” (sección 4.3.3.2) y “Conceptualización del Problema de Explotación de Información” (sección 4.3.3.3).

Se describe detalladamente la última fase *determinación del proceso de explotación de información* (sección 4.3.4), conformada por el paso: “Derivación del Proceso de Explotación de Información” (sección 4.3.4.1). En la sección 4.3.5 se brinda un cuadro que sintetiza los pasos a realizar en el *proceso de derivación de modelos*, indicando sus entradas y salidas.

4.3.1. Aspectos Generales

Luego de finalizado las primeras tres actividades del modelo de procesos para proyectos de explotación de información [Vanrell, 2011] (entendimiento del negocio, entendimiento de los datos y preparación de los datos), la primera destinada al entendimiento de los objetivos y requerimientos del proyecto desde una perspectiva del negocio, y las restantes a identificar, seleccionar y preparar los datos que son de utilidad para el proyecto, se poseen los elementos suficientes para dar comienzo al proceso de derivación de modelos, estos son: la descripción del dominio de negocio, la descripción de los datos del negocio, la descripción del problema de explotación de información y la descripción de los datos del problema de negocio. A partir de estos elementos es posible realizar las primeras dos fases del proceso propuesto (*análisis del dominio del negocio* y *análisis del problema de explotación de información*) debido a su independencia de conceptos, como se explicó previamente en el capítulo.

A los fines de favorecer la comprensión del *proceso de derivación de modelos*, las actividades serán desarrolladas en el orden propuesto, divididas en tres secciones según la fase a la que pertenecen (fase *orientada al análisis del dominio del negocio* (sección 4.3.2), fase *orientada al análisis del problema de explotación de información* (sección 4.3.3) y la fase *orientada a determinar el proceso de explotación de información* (sección 4.3.4)).

A efectos de ilustrar los entregables producidos en las distintas etapas del proceso propuesto, se presenta el siguiente caso de estudio [Basso y Vegega, 2013], con el fin de facilitar la comprensión de las tareas a realizar en cada uno de los pasos y la aplicación de las distintas técnicas utilizadas en ellos.

“...Descripción del Negocio y Problema de Negocio:

“Servicios Móviles Argentina S.A.” es una empresa perteneciente al sector de las telecomunicaciones que opera en todo el país. Ofrece productos de diversas marcas comerciales y genéricas, servicios y una amplia variedad de planes tarifarios (de uso personal o corporativo) de acuerdo a las necesidades de los clientes. Actualmente, está impulsando campañas de retención de clientes personalizadas en todo el país con el objetivo de ofrecer

nuevos productos y servicios acorde a las características de cada cliente, para adelantarse a sus competidores que también están creciendo a nivel nacional. El objetivo de negocio es caracterizar a los clientes en las diferentes regiones geográficas de cobertura de la empresa, con el fin de facilitar la definición de campañas de marketing orientadas al mantenimiento de los clientes actuales y a la obtención de nuevos clientes según las preferencias comunes de cada región, lo que ayudará también a mejorar la rentabilidad de la empresa. Con esto, se busca también determinar el comportamiento de sus clientes y mejorar el conocimiento que se tiene de ellos. “Servicios Móviles Argentina S.A.” tiene como meta la satisfacción de sus clientes, lo cual realiza a través del ofrecimiento de productos y servicios de calidad, dividiendo todas las operaciones de telefonía móvil del país en cinco regiones: Litoral, Cuyo, Pampeana/Centro, Patagonia y Norte.

Problema de explotación de información:

Determinar las reglas, en base a las variables situación, el tipo de plan y su duración, la marca del equipo y el tipo de cliente, que identifican el comportamiento de un conjunto de clientes de acuerdo a su ubicación geográfica (región, provincia, localidad).

Descripción de los datos:

Los siguientes son los datos relevantes almacenados por la empresa durante el ejercicio de sus actividades:

Código del Cliente: Código de cliente que la empresa utiliza para identificarlo. Ejemplo: 99861598.

Situación del Cliente: Situación del Cliente. Ejemplo: AAA (cliente activo), BA (cliente de baja).

Fecha de Alta: Fecha de alta del cliente. Ejemplo: 10/03/2003.

Código de la localidad: Código de localidad de residencia del cliente. Ejemplo: 379.

Código de la provincia: Código de la provincia a la cual pertenece la localidad. Ejemplo: 3.

Plan Contratado: Código de plan tarifario contratado por el cliente. Ejemplo: U21.

Duración: Cantidad de meses de contratación del cliente. Ejemplo: 12.

Tipo de Cliente: Tipo de cliente. Ejemplo: PR/PO (cliente prepago), FF (cliente con facturación).

Marca del Equipo: Marca del celular adquirido por el cliente. Ejemplo: NOKIA.

El conjunto de clientes será identificado mediante un valor numérico.

Región geográfica: Indica la región a la que pertenece cada provincia (tabla 4.2)...

REGIÓN	DESCRIPCIÓN	PROVINCIAS INTEGRANTES
1	Litoral	Misiones, Corrientes, Formosa, Chaco, Santa Fe
2	Cuyo	Mendoza, San Luis, San Juan
3	Pampeana/Centro	Buenos Aires, Capital Federal, Córdoba, La Pampa
4	Patagonia	Neuquén, Río Negro, Chubut, Santa Cruz, Tierra del Fuego
5	Norte	Jujuy, Salta, Tucumán, Catamarca, La Rioja, Santiago del Estero

Tabla 4.2. Código de regiones por provincias

4.3.2. Fase Orientada al Análisis del Dominio del Negocio

En esta sección se aborda detalladamente la primera fase del proceso propuesto, la cual abarca las actividades que solo se ejecutarán por única vez, independientemente de que se requiera del análisis de más de un problema de explotación de información. Esta fase tienen como objetivo la comprensión, definición y modelado del dominio del negocio, analizado en la primera etapa del ciclo de vida del proyecto, generándose mediante la aplicación de los pasos propuestos una serie de resultados intermedios (modelos de representación intermedia), a través de los cuales se logra obtener la representación final de la fase (pero intermedia desde el punto de vista del proceso) la Red Semántica del Modelo de Negocio. Los pasos que componen dicha fase son: *identificación de los elementos y estructura del dominio* (sección 4.3.2.1), *identificación de relaciones entre conceptos del dominio* (sección 4.3.2.2) y *conceptualización del dominio* (sección 4.3.2.3).

4.3.2.1. Paso 1: “Identificación de los elementos y estructura del dominio”

El paso “Identificación de los elementos y estructura del dominio”, tiene como objetivo generar a partir de la descripción del dominio de negocio y de los datos del negocio (elementos de entrada del paso), una estructura que defina los elementos que integran el modelo de negocio, en la que se identifiquen los conceptos que forman parte del mismo, junto con las características que lo definen.

En base a lo establecido previamente, es sustancial definir los siguientes términos: “conceptos” son aquellos elementos que poseen características las cuales permiten distinguirse de otros, que pueden interactuar con otros elementos, o que pueden ser agrupados o generalizados. Las características de los conceptos se las denomina “atributos”, los cuales pueden tener distintos “valores”, entendiéndose por estos el estado en el que se encuentra un atributo, y los que definen los distintos estados posibles que un atributo puede tener. Las “relaciones” son las interacciones entre los distintos elementos.

Los elementos de entrada de dicho paso son la descripción del dominio del negocio y los datos del negocio.

El desarrollo del paso consta de aplicar dos técnicas, la primera de ellas la tabla Término-Categoría-Definición del Dominio (TCDD), cuyo objetivo es identificar y clasificar todos los elementos relevantes pertenecientes al dominio del negocio (sección 4.3.2.1.1) y la segunda, la tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD), mediante la cual se define la estructura de los elementos del negocio (sección 4.3.2.1.2). Los productos resultados de aplicar ambas técnicas son elementos de salida del paso, los cuales serán utilizados posteriormente.

4.3.2.1.1. Técnica Tabla Término-Categoría-Definición del Dominio

La tabla Término-Categoría-Definición del Dominio del negocio, se basa en el formalismo tabla CCD propuesto en [Rodríguez, 2012].

Por medio de esta técnica, se identifican los elementos relevantes pertenecientes al modelo de negocio, posteriormente utilizados en otros formalismos orientados a dicho dominio. Los términos detectados serán introducidos en la tabla en orden alfabético. Un término pertenece a alguna de las siguientes categorías: Concepto, Atributo ó Relación (definidas en la sección previa). Para la ejecución de la tabla TCDD, se requiere como elemento de entrada a la descripción del dominio del negocio.

Para su aplicación, se procede a identificar en la descripción del dominio del negocio, aquellos elementos relevantes para la comprensión del mismo, a partir de las definiciones anteriormente

provistas. Por cada término detectado, se procederá con la identificación de la categoría a la que pertenece, y se definirá el mismo en base al dominio del negocio analizado.

La tabla 4.3 resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

Técnica Tabla Término-Categoría-Definición del Dominio

Entrada	Descripción del dominio del negocio
Salida	Tabla Término-Categoría-Definición del Dominio
Procedimiento	Paso 1. Identificar en el dominio del negocio aquellos términos relevantes para la comprensión del mismo Paso 2. Identificar la categoría a la que pertenece cada término identificado Paso 3. Definir cada término en base al dominio del negocio analizado

Tabla 4.3. Técnica Tabla Término-Categoría-Definición del Dominio

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la tabla 4.4. En esta tabla se observan todos los términos identificados, a partir de la descripción del dominio del negocio descripta en el caso de estudio brindado en la sección 4.3.1, luego se define la categoría a la que pertenecen, de acuerdo a las definiciones provistas en la sección 4.3.2.1, y se brinda una descripción, la cual define al termino en base al dominio de negocio analizado.

Término	Categoría	Definición
Cliente	Concepto	Persona que contrata el servicio
Código Cliente	Atributo	Identificación univoca de cada cliente
Código Plan	Atributo	Identificación univoca de cada plan
Contrata	Relación	El cliente contrata un plan determinado
Duración	Atributo	Cantidad de meses de contratación
es	Relación	El cliente es de un cierto tipo de clientes
Fecha alta	Atributo	Fecha en que se dio de alta en la empresa
identifica	Relación	El código de cliente identifica un cliente particular
identifica	Relación	El código de plan identifica a un plan particular
Localidad	Atributo	Localidad en la que reside el cliente
Marca equipo	Atributo	Marca de celular que el cliente posee.
pertenece	Relación	El cliente pertenece a una región
pertenece	Relación	El cliente pertenece a una provincia
pertenece	Relación	El cliente pertenece a una localidad
Plan	Concepto	Tipos de contratos que definen el servicio prestado
Provincia	Atributo	Provincia en la que reside el cliente
Región	Atributo	Región en la que reside el cliente
se registro	Relación	El cliente se registró en una fecha particular
Situación	Atributo	Estado en el que pertenece el cliente con la empresa
tiene	Relación	El cliente tiene una situación con respecto a la empresa
tiene	Relación	El cliente tiene un equipo telefónico de cierta marca
tiene	Relación	El plan tiene una duración determinada
Tipo cliente	Atributo	Clasificación del cliente según su facturación

Tabla 4.4. Tabla Término-Categoría-Definición del Dominio (TCDD) a partir del caso de estudio

4.3.2.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio

Por medio de esta técnica, se define la estructura del dominio de negocio, proporcionando un listado con los conceptos que se manipulan en el mismo. Cada concepto quedará descrito a través de los atributos que lo componen, el identificador que describe su relación con el concepto y los valores posibles de cada atributo. Para la ejecución de la tabla CARVD, se dispone como productos de entrada la tabla TCDD y la descripción de los datos del negocio.

Para su aplicación, se procede a identificar en la tabla TCDD, aquellos términos cuya categoría es “Concepto”, introduciéndolos en la columna etiquetada con el mismo nombre. Posteriormente se agrupan en la columna *atributo*, aquellos términos presentes en la tabla TCDD cuya categoría es “Atributo”, asignándolos en la fila correspondiente al concepto con el cual dicho atributo se relaciona, es decir, existe un término en la tabla TCDD, cuya categoría es “Relación”, que vincula al concepto con el atributo a asignar. El identificador del término de categoría “Relación” previamente identificado, será asignado en la columna *relación*, obteniéndose la estructura general del negocio. A partir de la descripción de los datos, se procede a identificar los posibles valores de cada atributo en la columna *valor*.

La tabla 4.5 resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio	
Entrada	Tabla Término-Categoría-Definición del Dominio Descripción de los datos del negocio
Salida	Tabla Concepto-Atributo-Relación-Valor del Dominio
Procedimiento	<p>Paso 1. identificar en la tabla TCDD, aquellos términos cuya categoría es "Concepto" y registrarlos en la columna homónima</p> <p>Paso 2. identificar en la tabla TCDD, aquellos términos cuya categoría es "Atributo" y se relacionen con uno de los conceptos previamente identificados, registrándolos en la columna <i>atributo</i> de la fila perteneciente al concepto asociado</p> <p>Paso 3. identificar en la tabla TCDD, el término "Relación" que vincula al concepto con el atributo que lo compone y registrar su nombre en la columna <i>relación</i></p> <p>Paso 4. registrar en base a la descripción de los datos del negocio, los valores posibles de cada atributo identificado</p>

Tabla 4.5. Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la tabla 4.6:

Concepto	Atributo	Relación	Valor
Cliente	Código	identifica	Numérico
	situación	tiene	Alfabético EJ: AAA, BA.
	fecha Alta	se registró	dd/mm/AAAA
	provincia	pertenece	Numérico EJ: 379
	localidad	pertenece	Numérico EJ: 3
	región	pertenece	1 al 5
	tipo cliente	es	Alfabético EJ: PR/PO, FF.
	marca equipo	tiene	Alfabético EJ: Nokia
Plan	Código	identifica	Alfanumérico EJ: U21.
	duración	tiene	Numérico

Tabla 4.6. Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) a partir del caso de estudio

La tabla 4.6, se obtiene al asignar en la columna *concepto* aquellos términos identificados en la tabla 4.4, cuya categoría es “Concepto”. Luego se asigna en la columna *atributo* aquellos términos cuya categoría es homónima, en la fila correspondiente al concepto con el cual se relaciona, y se registra la descripción del término de categoría “Relación” que vincula al concepto con el atributo en la columna *relación*. Finalmente, en la columna *valor* se asignan los posibles valores del atributo, de acuerdo a lo definido en la descripción de los datos (sección 4.3.1).

4.3.2.2. Paso 2: “Identificación de Relaciones entre Conceptos del Dominio”

El objetivo del paso “Identificación de Relaciones entre Conceptos del Dominio”, es detectar a partir de los conceptos identificados en el paso anterior, las relaciones existentes en el modelo del negocio, entre dichos conceptos, estableciendo los vínculos y modos de interacción que tienen los distintos elementos que componen al dominio del negocio. Los elementos de entrada de dicho paso son las tablas TCDD y CARVD generadas en el paso anterior.

Para la aplicación de dicho paso, se implementa la técnica tabla Concepto-Relación del Dominio, mediante la cual se identifican las relaciones entre los conceptos detectados según el dominio del negocio (sección 4.3.2.2.1). El elemento resultante de aplicar la técnica recién mencionada, será la salida del paso actual.

4.3.2.2.1. Técnica Tabla Concepto-Relación del Dominio

Por medio de esta técnica, se definen las interacciones entre los conceptos del dominio de negocio. Dicha tabla proporciona una lista de las relaciones entre los conceptos que se definen en el dominio. Cada relación quedará definida a través de los conceptos que la conforman, el nombre de la relación y una descripción de la misma. Para la aplicación de la tabla CRD, se dispone como productos de entrada las tablas TCDD y CARVD.

Para comenzar a aplicar la tabla CRD, se procede a identificar aquellos términos cuya categoría sea “Relación”, presentes en la tabla TCDD que indiquen un vínculo entre dos conceptos del problema identificados en la tabla CARVD. Por cada elemento identificado se registrará en la columna *concepto*, aquel concepto que genera dicha relación, en la columna *concepto asociado* aquel concepto vinculado, en la columna *relación* el nombre que describe el tipo de vínculo y la descripción que defina en base al dominio del negocio la relación identificada, en la columna homónima. Estos últimos dos elementos presentes en la tabla TCDD.

La tabla 4.7 resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

Técnica Tabla Concepto-Relación del Dominio	
Entrada	Tabla Término-Categoría-Definición del Dominio Tabla Concepto-Atributo-Relación-Valor del Dominio
Salida	Tabla Concepto-Relación del Dominio
Procedimiento	<p>Paso 1. identificar en la tabla TCDD, aquellos términos cuya categoría es "Relación" que vinculen dos conceptos identificados en la tabla CARVD y registrar su nombre identificativo y su descripción en las columnas correspondientes</p> <p>Paso 2. identificar los conceptos que intervienen en cada relación registrada, y registrar al concepto que origina dicha relación y al concepto vinculado, en las columnas concepto y concepto asociado respectivamente</p>

Tabla 4.7. Técnica Tabla Concepto-Relación del Dominio

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la tabla 4.8:

Concepto	Concepto asociado	Relación	Descripción
cliente	plan	contrata	El cliente contrata un plan por un tiempo determinado

Tabla 4.8. Tabla Concepto-Relación del Dominio (CRD) a partir del caso de estudio

En la tabla 4.8, se asignaron aquellos términos identificados en la tabla TCDD cuya categoría es “Relación” y en su descripción se vinculan dos conceptos, identificando en la columna “concepto”, aquel que genera la relación y en la columna “concepto asociado”, el concepto vinculado.

4.3.2.3. Paso 3: “Conceptualización Del Dominio”

A través del paso “Conceptualización Del Dominio”, se obtiene un modelo que representa gráficamente toda la información relevante del dominio del negocio, enfocado en las relaciones

entre los elementos que componen al negocio. Los elementos de entrada de dicho paso son las tablas CARVD y CRD, esta última obtenida del paso anterior.

Para su aplicación, se implementa la Técnica Red Semántica del Modelo de Negocio (sección 4.3.2.3.1) mediante la cual se obtiene la representación del dominio del negocio, y cuyo formalismo resultante será el elemento de salida del paso.

4.3.2.3.1. Técnica Red Semántica del Modelo de Negocio

Por medio de esta técnica, se expresan las interacciones entre los elementos del modelo de negocio, las cuales representan el conocimiento del negocio, desde un enfoque orientado a las relaciones de sus componentes. La conceptualización obtenida de aplicar dicha técnica, proporciona una visión general, simple y precisa de los elementos del negocio, sus características y sus relaciones. La notación a utilizar es la definida en la sección 4.2. Para la aplicación de la RSMN, se dispone como productos de entrada la tabla CARVD, que contiene la estructura del negocio, y CRD, que comprende las relaciones entre los conceptos del negocio.

Para comenzar a aplicar la RSMN, se procede a asociar los elementos (conceptos, atributos, relaciones y valores) identificados en la tabla CARVD obteniendo las estructuras de los conceptos implicados. A partir de las estructuras obtenidas, se asocian aquellas que se relacionen, en base a la tabla CRD, constituyendo la red semántica del modelo de negocio.

La tabla 4.9 resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

Técnica Red Semántica del Modelo de Negocio	
Entrada	Tabla Concepto-Atributo-Relación-Valor del Dominio Tabla Concepto-Relación del Dominio
Salida	Red Semántica del Modelo de Negocio
Procedimiento	Paso 1. Asociar los elementos (conceptos, atributos, relaciones y valores) identificados en la tabla CARVD obteniendo las estructuras de los conceptos implicados Paso 2. A partir de las estructuras, se asocian aquellas que se identifica una relación en la tabla CRD, constituyendo la red semántica del modelo de negocio

Tabla 4.9. Técnica Red Semántica del Modelo de Negocio

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la figura 4.12. Para realizar dicha figura, se grafica mediante las notaciones propuestas en la sección 4.2, los elementos identificados en la tabla CARVD. Se representan los conceptos y posteriormente se los relaciona con los atributos que lo componen, en base a las relaciones definidas. Una vez definida la estructura, se vinculan aquellos conceptos relacionados en la tabla CRD.

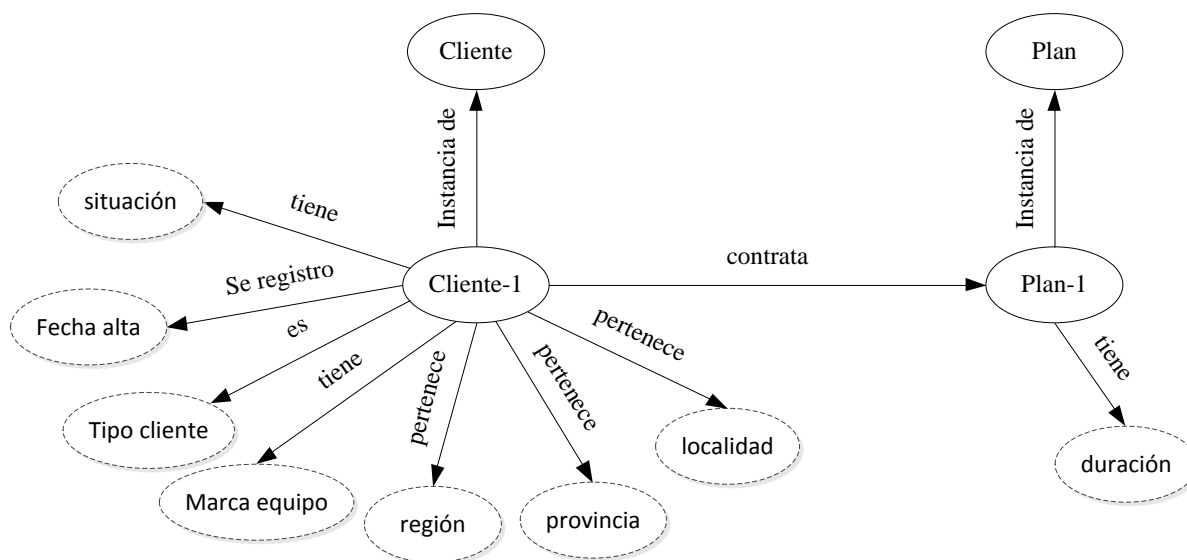


Figura 4.12. Red Semántica del Modelo de Negocio (RSMN) a partir del caso de estudio

4.3.3. Fase Orientada al Análisis del Problema de Explotación de Información

En esta sección se aborda detalladamente la segunda fase del proceso propuesto, la cual abarca aquellas actividades dedicadas a la comprensión y representación del problema de explotación de información, definido en la primera etapa del ciclo de vida del proyecto previamente mencionado.

La aplicación de esta fase puede realizarse en paralelo a la fase orientada al análisis del dominio del negocio, ya que esta tiene independencia con respecto a las salidas de las actividades de la fase previa. Esta fase deberá realizarse por cada problema de explotación de información a analizar.

A través de aplicar los pasos propuestos, se obtienen una serie de resultados intermedios (modelos de representación intermedia), mediante de los cuales se logra obtener la representación final de la fase (pero intermedia desde el punto de vista del proceso) la Red Semántica del Problema de Explotación de Información. Las etapas que la componen son: *identificación de los elementos y estructura del problema de explotación de información* (sección 4.3.3.1), *identificación de relaciones entre conceptos del problema de explotación de información* (sección 4.3.3.2) y *conceptualización del problema de explotación de información* (sección 4.3.3.3).

4.3.3.1. Paso 4: “Identificación de los elementos y estructura del Problema de Explotación de Información”

El objetivo del paso “Identificación de los elementos y estructura del Problema de Explotación de Información”, es detectar a partir de la descripción del problema de explotación de información los

distintos elementos que lo componen. A partir de los elementos detectados, junto con la descripción de los datos del problema de explotación de información se define la estructura de los conceptos identificados, se determinan las relaciones existentes entre conceptos y se define el flujo del problema de explotación de información mediante los elementos de entrada y salida del mismo, entendiéndose como los atributos a partir de los cuales se define el resultado y los atributos esperados como resultados respectivamente. Los elementos de entrada de dicho paso son la descripción del problema de explotación de información y los datos del problema de explotación de información.

El desarrollo del paso consta de aplicar dos técnicas, la tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI), cuyo objetivo es identificar y clasificar todos los elementos relevantes pertenecientes al problema de explotación de información (sección 4.3.3.1.1), y la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI), mediante la cual se define la estructura de los elementos que integran al problema y se define el flujo del problema, a partir de la identificación de sus elementos de entrada y salida (sección 4.3.3.1.2). Los productos resultados de aplicar estas técnicas son elementos de salida del paso, los cuales serán utilizados posteriormente.

4.3.3.1.1. Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información

La tabla Término-Categoría-Definición del Problema de Explotación de Información, se basa en el formalismo CCD propuesto en [Rodríguez, 2012]. Por medio de esta técnica, se identifican las ideas pertenecientes al problema de explotación de información, posteriormente utilizadas en formalismos orientados al problema de explotación. Los términos detectados serán introducidos en la tabla en orden alfabético. Un término puede ser categorizado en “Concepto”, “Atributo” o “Relación”, de acuerdo a lo definido en la sección 4.3.2.1. Para la ejecución de la tabla TCDPEI, se dispone como productos de entrada la descripción del problema de explotación de información y de los datos del problema de negocio.

Para comenzar a aplicar la tabla TCDPEI, se procede a identificar en el problema de explotación de información los elementos relevantes para la comprensión del mismo, a partir de las definiciones previamente descriptas, por cada término detectado, se procederá con la identificación de la categoría a la que pertenece, y se definirá el mismo en base al problema de explotación de información analizado.

La tabla 4.10 resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información	
Entrada	Descripción del problema de explotación de información Descripción de los datos del problema de explotación de información
Salida	Tabla Término-Categoría-Definición del Problema de Explotación de Información
Procedimiento	Paso 1. Identificar en el problema de explotación de información los elementos relevantes para la comprensión del mismo Paso 2. Por cada término detectado, identificar la categoría a la que pertenece cada elemento Paso 3. Definir cada elemento en base al problema de explotación de información analizado

Tabla 4.10. Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la tabla 4.11. En esta tabla se observan todos los términos identificados, a partir de la descripción del problema de explotación de información, descriptos en el caso de estudio brindado en la sección 4.3.1, luego se define la categoría a la que pertenecen, de acuerdo a las definiciones provistas en la sección 4.3.2.1, y se brinda una descripción que define a los mismos.

4.3.3.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información

Por medio de esta técnica, se define la estructura del problema de explotación de información, proporcionando un listado con los conceptos que se manipulan en dicho problema. Cada concepto quedará descripto a través de los atributos que lo componen, el identificador que describe su relación con el concepto y los valores posibles de cada atributo. La tabla Concepto-Atributo-Relación-Valor Extendida, presenta una columna adicional denominada *entrada/salida*, en la cual se identifican los atributos que actúan como factores iniciales y resultantes del problema de explotación de información identificándolos con la palabra “entrada” y “salida” respectivamente. Para la ejecución de la tabla CARVEPEI, se dispone como productos de entrada la tabla TCDPEI y la descripción de los datos y del problema de explotación de información.

Para la aplicación la tabla CARVEPEI, se identifican en la tabla TCDPEI, aquellos términos cuya categoría es “concepto”, introduciéndolos en la columna etiquetada con el mismo nombre.

Posteriormente se agrupan en la columna *atributo*, aquellos términos presentes en la tabla TCDPEI cuya categoría es “Atributo”, asignándolos en la fila correspondiente al concepto con el cual dicho atributo se relaciona, es decir, existe un término en la tabla TCDPEI, cuya categoría es “Relación”, que vincula al concepto con el atributo a asignar.

Término	Categoría	Definición
Cliente	Concepto	Persona que contrata el servicio
Código de grupo	Atributo	Identificación univoca de cada grupo
Código Plan	Atributo	Identificación univoca de cada plan
Código regla	Atributo	Identificación univoca de la regla
define	Relación	La regla define la asignación del cliente a un grupo determinado
duración	Atributo	Cantidad de meses de contratación
Grupo	Concepto	Conjunto de clientes
identifica	Relación	El código de plan identifica a un plan particular
identifica	Relación	El código de grupo identifica un grupo particular
identifica	Relación	El código regla identifica a una regla particular
integrado por	Relación	El grupo está integrado por clientes
Localidad	Atributo	Localidad en la que reside el cliente
Marca equipo	Atributo	Marca de celular que el cliente posee.
Plan	Concepto	Tipos de contratos que definen el servicio prestado
Provincia	Atributo	Provincia en la que reside el cliente
Región	Atributo	Región en la que reside el cliente
Regla	Concepto	Variables que definen la asignación de un cliente a un determinado grupo
según	Relación	El cliente pertenece a un grupo según su localidad
según	Relación	El cliente pertenece a un grupo según su provincia
según	Relación	El cliente pertenece a un grupo según su región
Situación	Atributo	Estado en el que se encuentra el cliente con la empresa
Subconjunto	Relación	La situación puede ser una variable que defina la asignación del cliente en un grupo
Subconjunto	Relación	El plan puede ser una variable que defina la asignación del cliente en un grupo
Subconjunto	Relación	La marca del celular puede ser una variable que defina la asignación del cliente en un grupo
Subconjunto	Relación	El tipo de cliente puede ser una variable que defina la asignación del cliente en un grupo
tiene	Relación	El plan tiene una duración determinada
Tipo cliente	Atributo	Clasificación del cliente según su facturación

Tabla 4.11. Tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) a partir del caso de estudio

El identificador del término de categoría “Relación” previamente identificado, será asignado en la columna *relación*, obteniéndose la estructura general del negocio. A partir de la descripción de los datos, se procede a identificar los posibles valores de cada atributo en la columna *valor*. Finalmente, se identifica en la columna *entrada/salida*, los atributos que actúan como entrada y salida del problema de explotación de información.

La tabla 4.12 resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la tabla 4.13.

Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información

Entrada	Descripción del problema de explotación de información Descripción de los datos del problema de explotación de información Tabla Término-Categoría-Definición del Problema de Explotación de Información
Salida	Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información
Procedimiento	<p>Paso 1. identificar en la tabla TCDPEI, aquellos términos cuya categoría es "Concepto" y registrarlos en la columna homónima</p> <p>Paso 2. identificar en la tabla TCDPEI, aquellos términos cuya categoría es "Atributo" y se relacionen con uno de los conceptos previamente identificados, registrándolos en la columna atributo de la fila perteneciente al concepto asociado</p> <p>Paso 3. identificar en la tabla TCDPEI, el término "Relación" que vincula al concepto con el atributo que lo compone y registrar su nombre en la columna relación</p> <p>Paso 4. registrar en base a la descripción de los datos del negocio, los valores posibles de cada atributo identificado</p> <p>Paso 5. registrar los atributos de entrada y salida del problema analizado</p>

Tabla 4.12. Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información

Concepto	Atributo	Relación	Entrada/Salida	Valor
Regla	Código regla	identifica	Salida	N Numérico
	Situación	subconjunto		A Alfabético EJ: AAA, BA.
	Marca equipo	subconjunto		A Alfabético EJ: Nokia
	Tipo cliente	subconjunto		A Alfabético EJ: PR/PO, FF.
Plan	Código Plan	identifica		A Alfanumérico EJ: U21.
	duración	tiene		N Numérico
Grupo	Código de grupo	identifica	Salida	N Numérico
Cliente	provincia	según	Entrada	N Numérico EJ: 379
	localidad	según	Entrada	N Numérico EJ: 3
	región	según	Entrada	1 al 5

Tabla 4.13. Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) a partir del caso de estudio

La tabla 4.13, se obtiene al asignar en la columna *concepto* aquellos términos identificados en la tabla 4.11, cuya categoría es “Concepto”. Luego se asigna en la columna *atributo* aquellos términos cuya categoría es homónima, en la fila correspondiente al concepto con el cual se relaciona, y se registra la descripción del término de categoría “Relación” que vincula al concepto con el atributo en la columna *relación*. Posteriormente, en la columna *valor* se asignan los posibles valores del atributo, de acuerdo a lo definido en la descripción de los datos (sección 4.3.1). Finalmente, se identifican en la columna *entrada/salida*, aquellos atributos que actúan como elementos de ingreso y resultado en el problema analizado.

4.3.3.2. Paso 5: “Identificación de relaciones entre conceptos del Problema de Explotación de Información”

El objetivo del paso “Identificación de relaciones entre conceptos del Problema de Explotación de Información”, es detectar a partir de los conceptos identificados en el paso anterior, las relaciones existentes en el problema de explotación de información, entre dichos conceptos, estableciendo los vínculos y modos de interacción que tienen los distintos elementos que componen al problema. Los elementos de entrada de dicho paso son las tablas TCDPEI y CARVEPEI generadas en el paso anterior.

Para la aplicación de dicho paso, se implementa la técnica tabla Concepto-Relación del Problema del Explotación de Información, mediante la cual se identifican las relaciones entre los conceptos detectados según el problema de explotación de información (sección 4.3.3.2.1). El elemento resultante de aplicar la técnica recién mencionada, será el elemento salida del paso actual.

4.3.3.2.1. Técnica Tabla Concepto-Relación del Problema de Explotación de Información

Por medio de esta técnica, se definen las interacciones entre los conceptos del problema de explotación de información. Dicha tabla proporciona un listado de las relaciones entre los conceptos que se definen en el problema tratado. Cada relación quedará definida a través de los conceptos que la conforman, el nombre de la relación y una descripción de la misma. Para la aplicación de la tabla CRPEI, se dispone como productos de entrada las tablas TCDPEI y CARVEPEI.

Para comenzar a aplicar la tabla CRPEI, se procede a identificar aquellos términos cuya categoría sea “Relación”, presentes en la tabla TCDPEI que indiquen un vínculo entre dos conceptos del problema identificados en la tabla CARVEPEI. Por cada elemento identificado se registrará en la columna *concepto*, aquel concepto que genera dicha relación, en la columna *concepto asociado* aquel concepto vinculado, en las columnas *relación* y *descripción*, se registra el nombre que describe el tipo de vínculo y la definición del mismo en base al problema de explotación de información, presentes en la tabla TCDPEI.

La tabla 4.14 resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la tabla 4.15.

Técnica Tabla Concepto-Relación del Problema de Explotación de Información

Entrada	Tabla Término-Categoría-Definición del Problema de Explotación de Información Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información
Salida	Tabla Concepto-Relación del Problema de Explotación de Información
Procedimiento	<p>Paso 1. identificar en la tabla TCDPEI, aquellos términos cuya categoría es "Relación" que vinculen dos conceptos identificados en la tabla CARVEPEI y registrar su nombre identificativo y su descripción en las columnas correspondientes</p> <p>Paso 2. identificar los conceptos que intervienen en cada relación registrada, y registrar al concepto que origina dicha relación y al concepto vinculado, en las columnas <i>concepto</i> y <i>concepto asociado</i> respectivamente</p>

Tabla 4.14. Técnica Tabla Concepto-Relación del Problema de Explotación de Información

Concepto	Concepto asociado	Relación	Descripción
Plan	Regla	Subconjunto	El plan puede ser una variable que defina la asignación del cliente en un grupo
Regla	Grupo	define	La regla define la asignación del cliente a un grupo determinado
Grupo	Cliente	integrado por	El grupo está integrado por clientes

Tabla 4.15. Tabla Concepto-Relación del Problema de Explotación de Información (CRPEI) a partir del caso de estudio

En la tabla 4.15, se asignaron aquellos términos identificados en la tabla TCDPEI cuya categoría es “Relación” y en su descripción se vinculan dos conceptos, identificando en la columna “concepto”, aquel que genera la relación y en la columna “concepto asociado”, el concepto vinculado, asignando en la columna “Relación” un nombre que identifique a la misma.

4.3.3.3. Paso 6: “Conceptualización del Problema de Explotación de Información”

A través del paso “Conceptualización del Problema de Explotación de Información”, se obtiene un modelo que representa gráficamente toda la información relevante del problema de explotación de información, enfocado en las relaciones entre los elementos que lo componen. Los elementos de entrada de dicho paso son las tablas CARVEPEI y CRPEI, esta última obtenida en el paso anterior. Para su aplicación, se implementa la Técnica Red Semántica del Problema de Explotación de Información (sección 4.3.3.3.1) mediante la cual se obtiene la representación de la estructura del problema de explotación de información, su flujo y las variables de inferencia definidas a través de los elementos de entrada y salida, y cuyo formalismo resultante será el elemento de salida del paso.

4.3.3.3.1. Técnica Red Semántica del Problema de Explotación de Información

Por medio de esta técnica, se expresan las interacciones entre los elementos del problema de explotación de información, las cuales representan el conocimiento del problema, desde un enfoque orientado a las relaciones de sus componentes. La notación a utilizar es la definida en la sección 4.2. Para la aplicación de la RSPEI, se dispone como productos de entrada la tabla CARVEPEI, que contiene la estructura del problema de explotación de información, y CRPEI, que comprende las relaciones entre los conceptos del problema.

Para su aplicación, se procede a asociar los elementos (conceptos, atributos, relaciones y valores) identificados en la tabla CARVEPEI obteniendo las estructuras de los conceptos implicados. A partir de las estructuras obtenidas, se asocian aquellas relacionadas, en base a la tabla CRPEI. Finalmente, se identifican los elementos de entrada, salida y las variables de inferencia en base a las notaciones específicas de las redes semánticas del problema de explotación de información definidas en la sección 4.2.2, constituyendo la RSPEI.

La tabla 4.16 resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

Técnica Red Semántica del Problema de Explotación de Información	
Entrada	Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información Tabla Concepto-Relación del Problema de Explotación de Información
Salida	Red Semántica del Problema de Explotación de Información
Procedimiento	<p>Paso 1. Asociar los elementos (conceptos, atributos, relaciones y valores) identificados en la tabla CARVEPEI obteniendo las estructuras de los conceptos implicados</p> <p>Paso 2. Asociar aquellas estructuras que se relacionen, en base a tabla CRPEI</p> <p>Paso 3. Identificar los nodos de entrada, salida y variables</p>

Tabla 4.16. Técnica Red Semántica del Problema de Explotación de Información

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la figura 4.13.

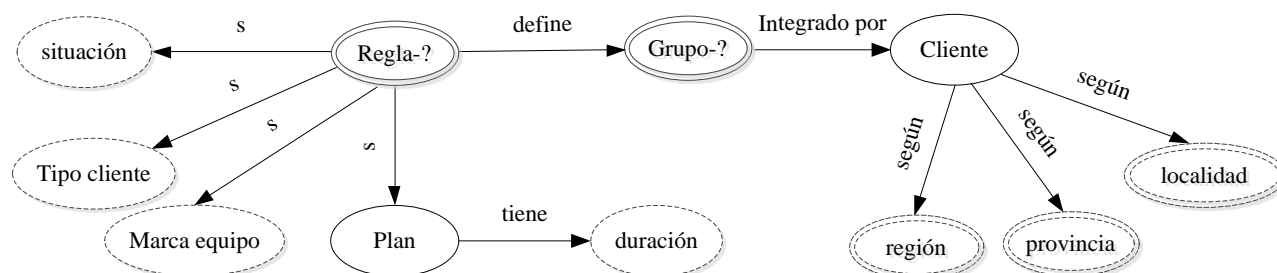


Figura 4.13. Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de estudio

Para realizar la figura 4.13, se grafica mediante las notaciones propuestas en la sección 4.2, los elementos identificados en la tabla CARVEPEI. Se representan los conceptos y posteriormente se los relaciona con los atributos que lo componen. Una vez definida la estructura, se vinculan aquellos conceptos relacionados en la tabla CRPEI. Los nodos variables identifican la incógnita definida en el problema de explotación, estos se ligan a los nodos constantes de la red (aquellos nodos que no son variables) para encontrar una equiparación que responda al problema planteado.

La interpretación del problema planteado es: *“determinar una regla compuesta por el subconjunto de los elementos situación, tipo de cliente, marca equipo, plan y la duración que tiene el plan, que definan al grupo (mediante su identificador) que está compuesto por clientes según su región, provincia y localidad.”*.

El significado de la interpretación debe implicar el mismo sentido que el problema de explotación definido a partir del problema de negocio presentado en la sección 4.3.1.

4.3.4. Fase de Determinación del Proceso de Explotación de Información

En esta sección se aborda detalladamente la última fase del proceso propuesto, cuyo objetivo es derivar a partir de las representaciones obtenidas en las fases anteriores, el proceso de explotación de información correspondiente.

En base a lo mencionado en el párrafo anterior, se puede inferir a partir de las dependencias que posee la fase, que para su ejecución es necesario haber finalizado las fases previas. Esta fase deberá realizarse por cada problema de explotación de información a analizar.

Esta fase se compone de la etapa *derivación del proceso de explotación de información* detallada a continuación (sección 4.3.4.1).

4.3.4.1. Paso 7: “Derivación del Proceso de Explotación de Información”

El objetivo del paso “derivación del proceso de explotación de información”, es brindar un mecanismo que permita deducir el proceso de explotación de información a aplicar para dar solución al problema de explotación de información definido. Para lograr dicho objetivo, se busca identificar en los modelos de representación intermedia, previamente realizados, un conjunto de patrones o estructuras comunes que permitan definir el proceso de explotación de información que provee la solución al problema. Los elementos de entrada son los resultados finales de cada fase previa la RSMN y la RSPEI.

Para su aplicación, se implementa la Técnica Algoritmo de Derivación del Proceso de Explotación de Información (sección 4.3.4.1.1), la cual define una serie de sub pasos que analizan la estructura

de las redes semánticas obtenidas como elemento de entrada, con el objetivo de definir el proceso de explotación de información, el cual es el resultado final del paso y del proceso propuesto.

4.3.4.1.1. Técnica Algoritmo de Derivación del Proceso de Explotación de Información

Por medio de esta técnica, se obtienen las técnicas y algoritmos de explotación de información que deben aplicarse para dar solución al problema de negocio planteado, pudiendo procesar la información disponible una vez que los datos se encuentren transformados y listos para ser utilizados, en otras palabras, el proceso de explotación de información. Este elemento resultante es la salida esperada del *proceso de derivación de modelos*. Para la aplicación del ADPEI, se dispone como productos de entrada la RSMN y la RSPEI.

Para comenzar a aplicar el ADPEI, se procede a identificar en las distintas redes semánticas ingresadas como elemento de entrada, la estructura descrita en el primer subpaso, variando la secuencia posterior de subpasos a ejecutar, a partir del resultado obtenido en cada subpaso. Como resultado de realizar los distintos caminos posibles (en base a las respuestas obtenidas en el caso analizado), se pueden obtener dos posibles respuestas:

- *Representación errónea*, ó
- El *proceso de explotación de información* correspondiente al problema definido.

La primera de las respuestas posibles, indica que alguna de las etapas previamente desarrolladas, se ha realizado de manera errónea. La segunda es la respuesta esperada y el resultado final del proceso. La tabla 4.17 resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

Técnica Algoritmo de Derivación del Proceso de Explotación de Información		
Entrada	Red Semántica del Modelo de Negocio	
	Red Semántica del Problema de Explotación de Información	
Salida	Proceso de explotación de información	
Procedimiento	Paso 1.	Identificar en las redes semánticas de entrada, aquellos patrones definidos en el Algoritmo de Derivación del Proceso de Explotación de Información

Tabla 4.17. Técnica Algoritmo de Derivación del Proceso de Explotación de Información

El paso general indicado en la tabla 4.17, que define el procedimiento está compuesto de 8 subpasos condicionales, que a partir del resultado obtenido al ejecutar el primer subpaso, se define el flujo de pasos mediante los cuales se obtiene el proceso de explotación de información a aplicar.

Para comprender cada uno de los subpasos que componen al ADPEI, es necesario definir los siguientes conceptos:

- Nodo variable son aquellos elementos interrogantes definidos en el problema de explotación de información, identificados en la red semántica con el símbolo “?”.
- Al menos debe existir un nodo variable en la red semántica del problema de explotación de información, de otra forma, la representación de la misma es incorrecta.

La tabla 4.18 define cada uno de los pasos y su destino de acuerdo al resultado obtenido. En dicha tabla, se puede observar que el primer paso busca identificar en la Red Semántica del Problema de Explotación de Información (RSPEI), la cantidad de nodos variables existentes.

ALGORITMO DE DERIVACIÓN DEL PROCESO DE EXPLOTACIÓN DE INFORMACIÓN		
Subpaso 7.1:	SI	en la Red Semántica del Problema de Explotación de Información (RSPEI) puede identificarse un único nodo variable
	ENTONCES	<i>Ir a Subpaso 7.2</i>
	SINO	<i>Ir a Subpaso 7.5</i>
Subpaso 7.2:	SI	el objetivo de la Red Semántica del Problema de Explotación de Información (RSPEI) es el identificador del nodo variable cuya arista (la cual surge a partir del nodo variable) se identifica con las palabras: pertenece, integra, es, es miembro de, u otras palabras de similar significado
	ENTONCES	<i>Aplicar Proceso de Descubrimiento de Grupos</i>
	SINO	<i>Ir a Subpaso 7.3</i>
Subpaso 7.3:	SI	el nodo variable identificado en la Red Semántica del Problema de Explotación de Información (RSPEI) se identifica con alguna de las siguientes palabras: elementos claves, factores destacados, elementos relevantes u otras palabras de similar significado
	ENTONCES	<i>Aplicar Proceso de Ponderación de Interdependencia de atributos</i>
	SINO	<i>Ir a Subpaso 7.4</i>
Subpaso 7.4:	SI	el nodo variable identificado en la Red Semántica del Problema de Explotación de Información (RSPEI) se identifica con las palabras: reglas, normas, criterios u otras palabras de similar significado
	ENTONCES	<i>Aplicar Proceso de Descubrimiento de Reglas de Comportamiento.</i>
	SINO	<i>Representación errónea.</i>
Subpaso 7.5:	SI	en la Red Semántica del Problema de Explotación de Información (RSPEI) se identifica un nodo variable el cual posee una arista, la cual surge a partir de este, identificada con alguna de las siguientes palabras: integrado por, compuesto por, formado por, u otras palabras de similar significado y se relaciona con otro nodo variable siendo este el nodo destino de la relación
	ENTONCES	<i>Ir a Subpaso 7.6.</i>
	SINO	<i>Representación errónea.</i>
Subpaso 7.6:	SI	el nodo variable de origen que integra la relación identificada en el Subpaso anterior, se identifica con las palabras: elementos claves, factores destacados, elementos relevantes u otras palabras de similar significado
	ENTONCES	<i>Ir a Subpaso 7.7.</i>
	SINO	<i>Ir a Subpaso 7.8.</i>
Subpaso 7.7:	SI	el nodo variable de destino que integra la relación, entre los nodos variables, referida en el Subpaso 7.5, es un nodo existente en la Red Semántica del Modelo del Negocio (RSMN)
	ENTONCES	<i>Aplicar Proceso de Ponderación de Reglas de Pertenencia a Grupos (Grupos definidos).</i>
	SINO	<i>Aplicar Proceso de Ponderación de Reglas de Pertenencia a Grupos (Grupos no definidos).</i>
Subpaso 7.8:	SI	el nodo variable de origen que integra la relación, entre los nodos variables, referida en el Subpaso 7.5, se identifica con las palabras: reglas, normas, criterios u otras palabras de similar significado
	ENTONCES	<i>Aplicar Proceso de Descubrimiento de Reglas de Pertenencia a Grupos.</i>
	SINO	<i>Representación errónea.</i>

Tabla 4.18. Algoritmo de Derivación del Proceso de Explotación de Información (ADPEI)

Si solo se identifica un nodo variable, se procede con el subpaso 7.2, el cual busca identificar si el único nodo variable en la RSPEI, tiene un arista la cual se identifica con las palabras pertenece, integra, es, es miembro de, u otras palabras de similar significado. Si la respuesta al paso 7.2 es afirmativa, entonces el proceso a aplicar es el *proceso de descubrimiento de grupos*, sino se procede

con el subpaso 7.3, el cual busca identificar si el nodo variable previamente reconocido, se identifica con algunas de las siguientes palabras: elementos claves, factores destacados, elementos relevantes u otras palabras de similar significado. Si la respuesta es afirmativa, se define que el proceso a aplicar es el *proceso de ponderación de interdependencia de atributos*, sino se procede con el paso 7.4, el cual busca identificar si el nodo variable previamente reconocido, se identifica con algunas de las siguientes palabras: reglas, normas, criterios u otras palabras de similar significado. Si la respuesta es afirmativa, entonces se debe aplicar el proceso de descubrimiento de reglas de comportamiento, sino se determina que la representación no es correcta.

Si la respuesta del subpaso 7.1 fuese mayor a uno, se procede a ejecutar el subpaso 7.5 el cual busca identificar en la RSPEI si alguno de los nodos variables posee una arista (originada en este), identificada con alguna de las siguientes palabras: integrado por, compuesto por, formado por, u otras palabras de similar significado y dicho nodo variable, se relaciona con otro nodo variable siendo este (el primer nodo) el nodo destino de la relación. Si la respuesta es negativa, se determina que la representación no es correcta, sino se prosigue con el subpaso 7.6 identificando si el nodo variable de origen (aquel que surge la arista que lo relaciona con el nodo variable identificado en el subpaso previo), se identifica con alguna de las siguientes palabras: elementos claves, factores destacados, elementos relevantes u otras palabras de similar significado.

Si la respuesta al subpaso 7.6 es afirmativa, se procede con el subpaso 7.7, que busca identificar si el nodo variable de destino, identificado en el subpaso 7.5, existe en la Red Semántica del Modelo del Negocio (RSMN). Si es un nodo existente en la conceptualización del dominio, el proceso a aplicar es el *Proceso Ponderación de Reglas de Pertenencia a Grupos (Grupos definidos)*, si este no existiese, entonces se aplica el *Proceso Ponderación de Reglas de Pertenencia a Grupos (Grupos no definidos)*.

Si la respuesta al subpaso 7.6 es negativa, se procede con el subpaso 7.8, el cual busca identificar si el nodo variable de origen (aquel nodo de donde surge la arista que lo relaciona con otro nodo variable) identificado en el paso 7.5, esta etiquetado con alguna de las siguientes palabras: reglas, normas, criterios u otras palabras de similar significado. En caso positivo, el proceso a aplicar es el *Proceso Descubrimiento de Reglas de Pertenencia a Grupos*, de otra forma, las representaciones previamente realizadas no son correctas.

Como resultado de aplicar la técnica definida al caso de estudio propuesto, se obtiene la siguiente secuencia de pasos, en la cual se presentan las redes implicadas con los elementos que comprueban el resultado obtenido de cada etapa, resaltados mediante líneas punteadas (tanto nodos como arcos) y mediante subrayado para textos:

Subpaso 7.1: Se identifican dos nodos variables en la Red Semántica del Problema de Explotación de Información (figura 4.14).

✓ Resultado obtenido: Ir a Subpaso 7.5

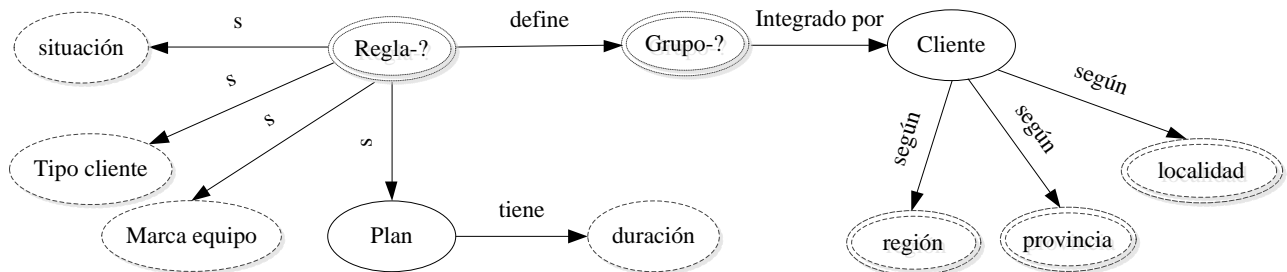


Figura 4.14. Identificación de subpaso 7.1 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de estudio

Subpaso 7.5: En la Red Semántica del Problema de Explotación de Información, uno de los nodos variable (etiquetado como “Grupo”), posee una arista identificada con las palabras “integrado por” y es el nodo destino de la relación “define” con el nodo variable “Regla” (figura 4.15).

✓ Resultado obtenido: Ir a Subpaso 7.6

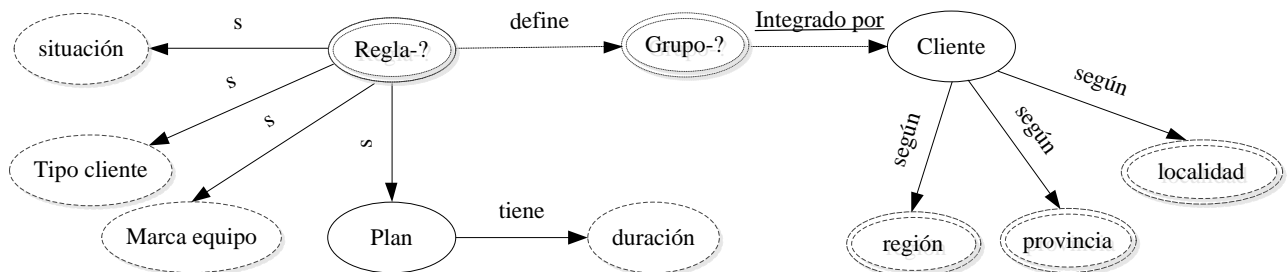


Figura 4.15. Identificación de subpaso 7.5 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de estudio

Subpaso 7.6: En la Red Semántica del Problema de Explotación de Información, no se identifica un nodo variable etiquetado con las palabras mencionadas.

✓ Resultado obtenido: Ir a Subpaso 7.8

Subpaso 7.8: En la Red Semántica del Problema de Explotación de Información, el nodo variable origen de la relación detectada en el subpaso 7.5, se identifica con la palabra “Regla” (figura 4.16).

✓ Resultado obtenido: **Aplicar Proceso de Descubrimiento de Reglas de Pertenencia a Grupos**

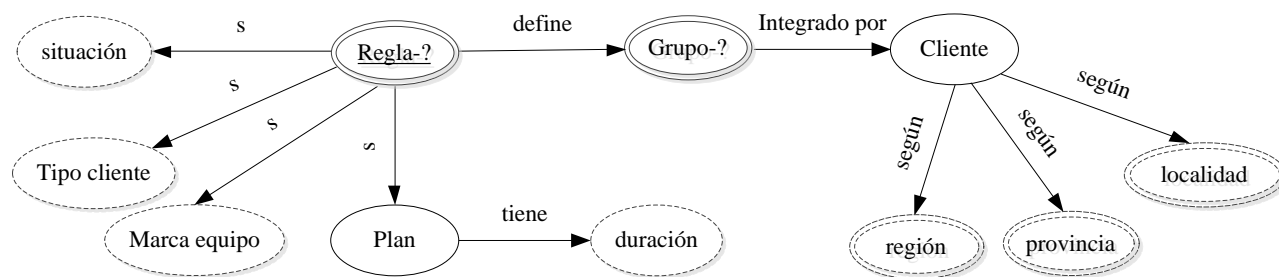


Figura 4.16. Identificación de subpaso 7.8 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de estudio

Retornando como respuesta del paso y del *proceso de derivación de modelos* aplicado al caso de estudio, el texto:

“Proceso de Descubrimiento de Reglas de Pertenencia a Grupos”.

4.3.5. Resumen del Proceso de Derivación de Modelos

En esta sección se presenta un resumen del *proceso de derivación de modelos* propuesto (tabla 4.19), que describe la secuencia de pasos que componen dicho proceso, junto con sus entradas y salidas correspondientes.

		INPUT		OUTPUT	
		ENTRADA	REPRESENTACIÓN	SALIDA	REPRESENTACIÓN
Paso 1:	Identificación de los elementos y estructura del dominio	<ul style="list-style-type: none"> Descripción del dominio del negocio Datos del negocio 	<ul style="list-style-type: none"> Texto plano 	<ul style="list-style-type: none"> Representación sistemática de los elementos relevantes del dominio Representación sistemática de la estructura general del dominio 	<ul style="list-style-type: none"> Tabla Término-Categoría-Definición del Dominio (TCDD) Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD)
Paso 2:	Identificación de relaciones entre conceptos del dominio	<ul style="list-style-type: none"> Representación sistemática de los elementos relevantes del dominio Representación sistemática de la estructura general del dominio 	<ul style="list-style-type: none"> Tabla Término-Categoría-Definición del Dominio (TCDD) Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) 	<ul style="list-style-type: none"> Representación sistemática de las relaciones entre los Conceptos del dominio 	<ul style="list-style-type: none"> Tabla Concepto-Relación del Dominio (CRD)
Paso 3:	Conceptualización del dominio	<ul style="list-style-type: none"> Representación sistemática de la estructura general del dominio Representación sistemática de las relaciones entre los Conceptos del dominio 	<ul style="list-style-type: none"> Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) Tabla Concepto-Relación del Dominio (CRD) 	<ul style="list-style-type: none"> Representación sistemática del modelo de negocio 	<ul style="list-style-type: none"> Red semántica del modelo de negocio (RSMN)
Paso 4:	Identificación de los elementos y estructura del Problema de Explotación de Información	<ul style="list-style-type: none"> Problema de explotación de información Datos del problema de explotación de información 	<ul style="list-style-type: none"> Texto plano 	<ul style="list-style-type: none"> Representación sistemática de los elementos relevantes del dominio Representación sistemática de la estructura del Problema de Explotación de Información y sus elementos de entrada y salida 	<ul style="list-style-type: none"> Tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI)

Tabla 4.19.a. Resumen del Proceso de Derivación de Modelos

		INPUT		OUTPUT	
		ENTRADA	REPRESENTACIÓN	SALIDA	REPRESENTACIÓN
Paso 5:	Identificación de relaciones entre conceptos del Problema de Explotación de Información	<ul style="list-style-type: none"> Representación sistemática de los elementos relevantes del dominio Representación sistemática de la estructura del Problema de Explotación de Información y sus elementos de entrada y salida 	<ul style="list-style-type: none"> Tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) 	<ul style="list-style-type: none"> Representación sistemática de las relaciones entre los Conceptos del Problema de Explotación de Información 	<ul style="list-style-type: none"> Tabla Concepto-Relación del Problema del Explotación de Información (CRPEI)
Paso 6:	Conceptualización del Problema de Explotación de Información	<ul style="list-style-type: none"> Representación sistemática de la estructura del Problema de Explotación de Información y sus elementos de entrada y salida Representación sistemática de las relaciones entre los Conceptos del Problema de Explotación de Información 	<ul style="list-style-type: none"> Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) Tabla Concepto-Relación del Problema del Explotación de Información (CRPEI) 	<ul style="list-style-type: none"> Representación sistemática del problema de explotación de información 	<ul style="list-style-type: none"> Red semántica del problema de explotación de información (RSPEI)
Paso 7:	Derivación del Proceso de Explotación de Información	<ul style="list-style-type: none"> Representación sistemática del modelo de negocio Representación sistemática del problema de explotación de información 	<ul style="list-style-type: none"> Red semántica del modelo de negocio (RSMN) Red semántica del problema de explotación de información (RSPEI) 	<ul style="list-style-type: none"> Proceso de explotación de información 	<ul style="list-style-type: none"> Texto plano

Tabla 4.19.b. Resumen del Proceso de Derivación de Modelos

5. CASOS DE VALIDACIÓN

En este capítulo se presentan tres casos de validación, pertenecientes a distintos dominios de conocimiento, los cuales presentan diferentes características para la aplicación del *proceso de derivación de modelos* propuesto y de las técnicas asociadas, a los efectos de implementar las tareas correspondientes a cada una de las fases. En la sección 5.1 se analiza un primer caso de validación correspondiente a una empresa de prestación de servicios de internet [Britos, 2008], la cual busca mejorar la permanencia de sus clientes. En la sección 5.2 se analiza un segundo caso de validación correspondiente a un negocio perteneciente a la industria automotriz [Flores, 2009], cuyo objetivo principal es la detección e identificación de daños y/o averías producidas a lo largo del circuito que recorren los autos desde que salen de la planta de producción hasta que llegan a su destino final. En la sección 5.3 se analiza el tercer caso de validación correspondiente a un análisis de la canasta de compras “market basket analysis”, en el cual se desea identificar las dependencias entre los productos [Giudici, 2003].

5.1. CASO DE VALIDACIÓN: EMPRESA DE PRESTACIÓN DE SERVICIOS DE INTERNET

En esta sección se analiza el primer caso de validación correspondiente a una empresa de prestación de servicios de internet. En la sección 5.1.1 se describe al modelo de negocio, sus datos y el problema de explotación de información a analizar. En la sección 5.1.2 se aplican las etapas y técnicas utilizadas para el desarrollo de las tareas correspondientes a la fase de *análisis del dominio del negocio*. En la sección 5.1.3 se implementan las etapas y técnicas utilizadas para el desarrollo de las tareas correspondientes a la fase de *análisis del problema de explotación de información*. En la sección 5.1.4 se aplica la etapa y su correspondiente técnica, utilizada para el desarrollo de las tareas correspondientes a la fase de *determinación del proceso de explotación de información* obteniendo como resultado el proceso de explotación de información a aplicar, en base al modelo del negocio y el problema analizado.

5.1.1. Descripción

En esta sección se presenta la información necesaria para comprender el modelo de negocio y el problema de explotación de información a analizar. El presente caso se obtiene de [Britos, 2008], y se presentan las características relevantes del mismo:

“...Descripción del Negocio y Problema de Negocio:

Una empresa de prestación de servicios busca comprender en mayor medida el comportamiento de sus clientes, con el fin de aumentar su permanencia en la empresa lo máximo posible. Para ello se cuenta con un conjunto de datos pertenecientes a una encuesta de satisfacción sobre el servicio realizada por la empresa.

La encuesta contempla los siguientes datos: la satisfacción general del cliente respecto del producto, si recomendaría el producto, si compraría nuevos productos a la compañía, si cambiaría el proveedor en los próximos 6 meses por una nueva propuesta, si el funcionamiento es el esperado, si considera que el gasto se corresponde con la calidad del servicio, cuanta competencia existe en la zona del cliente encuestado y si ha realizado algún movimiento en el producto en los últimos 6 meses (alta o baja de productos).

Se busca dar una descripción de las causales de alta o baja del servicio de dial-up de Internet provista por una compañía telefónica.

Problema de explotación de información:

Determinar las características que definen los movimientos del usuario.

Descripción de los datos:

Los datos que se poseen a partir de la encuesta son:

Satisfacción general: es un indicador de la valoración general que tiene el encuestado del producto. Sus posibles valores son: satisfecho ó insatisfecho.

Recomendación: indica si el cliente recomendaría el producto. Sus posibles valores son: satisfecho ó insatisfecho.

Continua comprando: indica la valoración del cliente respecto de continuar comprando productos de la compañía que tiene el encuestado: Sus posibles valores son: satisfecho ó insatisfecho.

Cambia proveedor: indica la valoración respecto de cambiar el proveedor del producto que tiene el encuestado. Sus posibles valores son: satisfecho ó insatisfecho.

Funcionamiento: indica la valoración respecto del funcionamiento del producto que tiene el encuestado. Sus posibles valores son: satisfecho ó insatisfecho.

Gasto: indica la valoración de la relación gasto/producto que tiene el encuestado. Sus posibles valores son: satisfecho ó insatisfecho.

Competencia: Indica la existencia de competidores en la zona del mismo producto. Sus posibles valores son: alta, media ó baja.

Movimientos: indicador de los movimientos realizados por el encuestado en los últimos 6 meses. Sus posibles valores son: alta, baja ó sin movimientos...”

5.1.2. Aplicación de las Técnicas Utilizadas en la Fase de Análisis del Dominio del Negocio

En esta sección se aplican, al caso de validación presente, las etapas y técnicas utilizadas para el desarrollo de las tareas correspondientes a la fase de análisis del dominio del negocio. Las etapas que componen a la fase actual son: *identificación de los elementos y estructura del dominio* (sección 5.1.2.1), *identificación de relaciones entre conceptos del dominio* (sección 5.1.2.2) y *conceptualización del dominio* (sección 5.1.2.3).

5.1.2.1. Paso 1: “Identificación de los elementos y estructura del dominio”

En este paso se implementan dos técnicas: la *tabla término-categoría-definición del dominio* (sección 5.1.2.1.1), orientada a identificar los elementos relevantes del modelo de negocio y la *tabla concepto-atributo-relación-valor del dominio* (sección 5.1.2.1.2), cuyo objetivo es definir la

estructura de los conceptos relevantes que integran al negocio. La descripción del dominio y los datos del negocio ingresan como elemento de entrada del paso, y las tablas término-categoría-definición del dominio (TCDD) y tabla concepto-atributo-relación-valor del dominio (CARVD), son los productos salida del paso.

5.1.2.1.1. Técnica Tabla Término-Categoría-Definición del Dominio

Esta técnica está compuesta por tres pasos, y genera como resultado de aplicar dichos pasos a la descripción del negocio provista en la sección 5.1.1, la tabla término-categoría-definición del dominio (TCDD).

Paso 1. *Identificación de términos relevantes:* se lleva a cabo el análisis de la descripción del negocio a los efectos de identificar y registrar alfabéticamente los elementos que permiten comprender el negocio.

A continuación se presenta el listado de términos identificados, ordenado alfabéticamente:

“cambia proveedor”, “Cliente”, “código cliente”, “competencia”, “continua comprando”, “describe”, “describe”, “funcionamiento”, “gasto”, “hay”, “identifica”, “movimientos”, “posee”, “posee”, “presenta”, “realiza”, “recomendación”, “satisfacción general” y “tiene”.

Paso 2. *Clasificación de términos por categoría:* a partir de los términos identificados en el paso anterior, se los clasifica de acuerdo al tipo de elemento que representa (sección 4.3.2.1).

A continuación se presenta el listado de términos identificados en el paso anterior, junto con la categoría asignada a cada uno de ellos, representando dicha asignación mediante pares ordenados, donde el primer elemento es el nombre del término y el segundo la categoría designada:

(cambia proveedor ; Atributo), (Cliente ; Concepto), (código cliente ; Atributo), (competencia ; Atributo), (continua comprando ; Atributo), (describe ; Relación), (describe ; Relación), (funcionamiento ; Atributo), (gasto ; Atributo), (hay ; Relación), (identifica ; Relación), (movimientos ; Atributo), (posee ; Relación), (posee ; Relación), (presenta ; Relación), (realiza ; Relación), (recomendación ; Atributo), (satisfacción general ; Atributo) y (tiene ; Relación).

Paso 3. *Describir el término:* se lleva a cabo una descripción de cada término identificado, en base al dominio del negocio que integra.

Los datos producidos en los pasos precedentes se integran a la tabla Término-Categoría-Definición del Dominio (TCDD) junto con la descripción de cada concepto. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.1 en la cual se observa el conjunto de términos identificados, su categoría y la descripción que los define de acuerdo al negocio analizado.

Término	Categoría	Definición
cambia proveedor	Atributo	Indica la valoración de si el cliente cambiaría de proveedor del servicio
Cliente	Concepto	Persona que contrata alguno de los servicios de la empresa
código cliente	Atributo	Identificador de cada cliente
competencia	Atributo	Indica la existencia de competidores del mismo producto en la zona
continúa comprando	Atributo	Indica la valoración de si el cliente continuaría comprando
describe	Relación	el cliente describe su satisfacción con el funcionamiento del servicio
describe	Relación	el cliente describe su satisfacción en base al gasto dado el funcionar del producto
funcionamiento	Atributo	Indica la valoración del cliente respecto el funcionamiento del servicio
gasto	Atributo	Indica la valoración entre el gasto y el producto por parte del cliente
hay	Relación	en la zona del cliente hay un nivel de competencia
identifica	Relación	el código de cliente identifica cada cliente
movimientos	Atributo	indica el movimiento realizado por el cliente en los últimos 6 meses
posee	Relación	el cliente posee una valoración de continuar comprando
posee	Relación	el cliente posee una valoración de cambio de proveedor
presenta	Relación	el cliente presenta una satisfacción general
realiza	Relación	el cliente realiza movimientos
recomendación	Atributo	Indica la valoración de si el cliente recomendaría el producto
satisfacción general	Atributo	Indica la valoración general del cliente sobre el producto
tiene	Relación	el cliente tiene una valoración de recomendación

Tabla 5.1. Tabla Término-Categoría-Definición del Dominio (TCDD) a partir del caso de validación: prestación de servicios de internet

5.1.2.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio

Esta técnica está compuesta por cuatro pasos, cuyos elementos de entrada son la descripción de los datos del negocio provista en la sección 5.1.1, y la tabla término-categoría-definición del dominio (TCDD) obtenida de ejecutar la técnica tabla término-categoría-definición del dominio. El producto resultado del paso es la tabla concepto-atributo-relación-valor del dominio (CARVD).

Paso 1. Identificación de conceptos: a partir de los términos identificados en la tabla término-categoría-definición del dominio (TCDD), se registran aquellos elementos de categoría “Concepto” en la columna homónima de la tabla concepto-atributo-relación-valor del dominio (CARVD).

Se identifica un único elemento de categoría “Concepto”: Cliente.

Paso 2. *Identificación de atributos por concepto:* se lleva a cabo una integración de cada término cuya categoría es “Atributo” identificado en la tabla término-categoría-definición del dominio (TCDD), al concepto previamente identificado en la tabla concepto-atributo-relación-valor del dominio (CARVD).

A continuación se presenta una lista de pares ordenados, en donde el primer elemento corresponde al concepto identificado en el paso previo, y el segundo al atributo asociado al concepto:

(Cliente ; código cliente), (Cliente ; satisfacción general), (Cliente ; recomendación), (Cliente ; continua comprando), (Cliente ; cambia proveedor), (Cliente ; funcionamiento), (Cliente ; gasto), (Cliente ; competencia) y (Cliente ; movimientos).

Paso 3. *Identificación de la relación:* se identifica y registra aquel elemento de la tabla término-categoría-definición del dominio (TCDD) cuya categoría es “Relación” que relaciona al concepto con el atributo, registrándolo en la columna homónima del atributo correspondiente.

A continuación se presenta una lista de ternas ordenadas, cuyos primeros dos elementos son los identificados en el punto anterior, agregándose el nombre de la relación que los vincula:

(Cliente ; código cliente ; identifica), (Cliente ; satisfacción general ; presenta), (Cliente ; recomendación ; tiene), (Cliente ; continua comprando ; posee), (Cliente ; cambia proveedor ; posee), (Cliente ; funcionamiento ; describe), (Cliente ; gasto ; describe), (Cliente ; competencia ; hay) y (Cliente ; movimientos ; realiza).

Paso 4. *Identificación de los valores posibles:* a partir de la descripción de los datos, se definen los posibles valores de cada atributo identificado.

Los datos producidos en los pasos precedentes se integran a la tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) junto con los valores posibles de cada atributo. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.2 en la cual se observa al conjunto de términos identificados en la técnica anterior, presentados de una forma organizada la cual permite describir la estructura de cada concepto perteneciente al dominio del negocio.

Concepto	Atributo	Relación	Valor
Cliente	código cliente	identifica	Numérico
	satisfacción general	presenta	Satisfecho/Insatisfecho
	recomendación	tiene	Satisfecho/Insatisfecho
	continúa comprando	posee	Satisfecho/Insatisfecho
	cambia proveedor	posee	Satisfecho/Insatisfecho
	funcionamiento	describe	Satisfecho/Insatisfecho
	gasto	describe	Satisfecho/Insatisfecho
	competencia	hay	Alta/Media/Baja
	movimientos	realiza	Alta/Baja/Sin movimientos

Tabla 5.2. Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) a partir del caso de validación: prestación de servicios de internet

5.1.2.2. Paso 2: “Identificación de Relaciones entre Conceptos del Dominio”

En este paso se implementan la técnica *tabla concepto-relación del dominio* (sección 5.1.2.2.1), orientada a identificar las interacciones entre los conceptos del dominio de negocio previamente identificados. Las tablas término-categoría-definición del dominio (TCDD) y concepto-atributo-relación-valor del dominio (CARVD), obtenidas a partir del paso 1, ingresan como elemento de entrada del paso actual, y la tabla concepto-relación del dominio (CRD) es su producto salida.

5.1.2.2.1. Técnica Tabla Concepto-Relación del Dominio

Esta técnica está compuesta por dos pasos, y genera como resultado de aplicar dichos pasos a las tablas término-categoría-definición del dominio (TCDD) y concepto-atributo-relación-valor del dominio (CARVD) obtenidas en la sección 5.1.2.1, la tabla concepto-relación del dominio (CRD).

Paso 1. *Identificación de relaciones entre conceptos:* se registran aquellos términos de categoría “Relación”, identificados en la tabla término-categoría-definición del dominio (TCDD), que relacionen a dos conceptos identificados, y se describe a dicha relación.

No se identifican términos que vinculen conceptos.

Paso 2. *Identificación ordenada de conceptos relacionados:* se registran los conceptos integrantes de las relaciones identificadas en el paso anterior, ingresando al concepto generador de la relación en la columna “Concepto” y al concepto asociado en la columna homónima.

El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.3 en la cual no se observan relaciones entre conceptos, dado que el negocio presenta un único concepto.

Concepto	Concepto asociado	Relación	Descripción

Tabla 5.3. Tabla Concepto-Relación del Dominio (CRD) a partir del caso de validación: prestación de servicios de internet

5.1.2.3. Paso 3: “Conceptualización Del Dominio”

En este paso se implementan la técnica *red semántica del modelo de negocio* (sección 5.1.2.3.1), cuyo objetivo es conceptualizar la información del dominio del negocio, desde un enfoque orientado a las relaciones de sus componentes. Las tablas concepto-atributo-relación-valor del dominio (CARVD) y la tabla concepto-relación del dominio (CRD), obtenidas en los pasos 1 y 2 respectivamente, ingresan como elemento de entrada del paso actual, y la red semántica del modelo de negocio (RSMN) es su producto salida.

5.1.2.3.1. Técnica Red Semántica del Modelo de Negocio

Esta técnica está compuesta por dos pasos, y genera como resultado de aplicar dichos pasos a las tablas concepto-atributo-relación-valor del dominio (CARVD) y concepto-relación del dominio (CRD) obtenidas en la sección 5.1.2.1.2 y 5.1.2.2.1 respectivamente, la red semántica del modelo de negocio (RSMN).

Paso 1. Representación de los conceptos: se representa a partir de la notación definida en la sección 4.2, la estructura de los conceptos identificados en la tabla concepto-atributo-relación-valor del dominio (CARVD).

El producto obtenido de aplicar dicho paso al presente caso de estudio, se ilustra en la figura 5.1 en la cual se observa al modelo de negocio descrito a través de las notaciones definidas, desde una perspectiva orientada a las relaciones entre componentes.

Paso 2. Representación de las interacciones entre conceptos: se representan las relaciones entre conceptos identificadas en la tabla concepto-relación del dominio (CRD), mediante la anotación provista en la sección 4.2.

Dado que no se identificaron relaciones entre conceptos, este paso no realiza modificaciones sobre el resultado obtenido en el paso previo, siendo el resultado final de la técnica la figura 5.1.

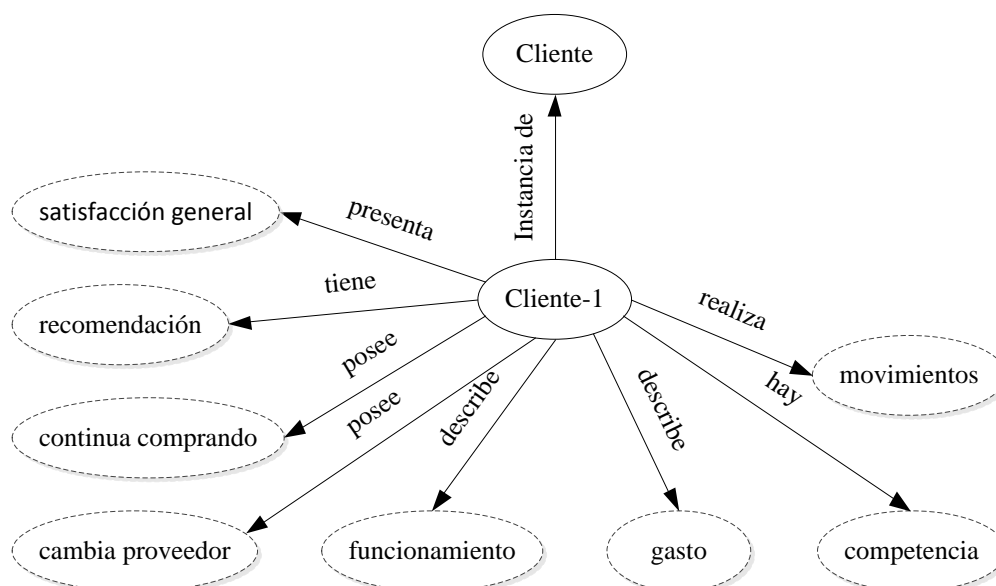


Figura 5.1. Red Semántica del Modelo de Negocio (RSMN) a partir del caso de validación: prestación de servicios de internet

5.1.3. Aplicación de las Técnicas Utilizadas en la Fase de Análisis del Problema de Explotación de Información

En esta sección se aplican al presente caso de validación las etapas y técnicas utilizadas para el desarrollo de las tareas correspondientes a la fase de análisis del problema de explotación de información. Las etapas que componen la fase son: *identificación de los elementos y estructura del problema de explotación de información* (sección 5.1.3.1), *identificación de relaciones entre conceptos del problema de explotación de información* (sección 5.1.3.2) y *conceptualización del problema de explotación de información* (sección 5.1.3.3).

5.1.3.1. Paso 4: “Identificación de los elementos y estructura del Problema de Explotación de Información”

En este paso se implementan dos técnicas: la *tabla término-categoría-definición del problema de explotación de información* (sección 5.1.3.1.1), orientada a identificar los elementos relevantes del problema de explotación de información y la *tabla concepto-atributo-relación-valor del problema de explotación de información* (sección 5.1.3.1.2), cuyo objetivo es definir la estructura de los conceptos relevantes que integran al problema. La descripción y los datos del problema ingresan como elemento de entrada del paso, y las tablas término-categoría-definición del problema de explotación de información (TCDPEI) y tabla concepto-atributo-relación-valor extendida del problema de explotación de información (CARVEPEI), son los productos salida del paso.

5.1.3.1.1. Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información

Esta técnica está compuesta por tres pasos, y genera como resultado de aplicar dichos pasos a la descripción del problema de explotación de información provista en la sección 5.1.1, la tabla término-categoría-definición del problema de explotación de información (TCDPEI).

Paso 1. *Identificación de términos relevantes:* se lleva a cabo el análisis de la descripción del problema de explotación de información a los efectos de identificar y registrar alfabéticamente los elementos que permiten comprender al mismo.

A continuación se presenta el listado de términos identificados, ordenado alfabéticamente:

“cambia proveedor”, “Cliente”, “código cliente”, “código regla”, “competencia”, “continua comprando”, “define”, “funcionamiento”, “gasto”, “identifica”, “identifica”, “movimientos”, “realiza”, “recomendación”, “Regla”, “satisfacción general”, “Subconjunto”, “Subconjunto”, “Subconjunto”, “Subconjunto”, “Subconjunto” y “Subconjunto”.

Paso 2. *Clasificación de términos por categoría:* a partir de los términos identificados en el paso anterior, se los clasifica de acuerdo al tipo de elemento que representa (sección 4.3.2.1).

A continuación se presenta el listado de términos identificados en el paso anterior, junto con la categoría asignada a cada uno de ellos, representando dicha asignación mediante pares ordenados, donde el primer elemento es el nombre del término y el segundo la categoría designada:

(cambia proveedor ; Atributo), (Cliente ; Concepto), (código cliente ; Atributo), (código regla ; Atributo), (competencia ; Atributo), (continua comprando ; Atributo), (define ; Relación), (funcionamiento ; Atributo), (gasto ; Atributo), (identifica ; Relación), (identifica ; Relación), (movimientos ; Atributo), (realiza ; Relación), (recomendación ; Atributo), (Regla ; Concepto), (satisfacción general ; Atributo), (Subconjunto ; Relación), (Subconjunto ; Relación), (Subconjunto ; Relación), (Subconjunto ; Relación), (Subconjunto ; Relación) y (Subconjunto ; Relación).

Paso 3. *Describir el término:* se lleva a cabo una descripción de cada término identificado, en base al problema de explotación de información que integra.

Los datos producidos en los pasos precedentes se integran a la tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) junto con la descripción de cada concepto. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.4 en la cual se observa el conjunto de términos identificados, su categoría y la descripción que los define en base al problema de explotación de información analizado.

Término	Categoría	Definición
cambia proveedor	Atributo	Indica la valoración de si el cliente cambiaría de proveedor del servicio
Cliente	Concepto	Persona que contrata alguno de los servicios de la empresa
código cliente	Atributo	Identificador de cada cliente
Código regla	Atributo	Identificación univoca de la regla
competencia	Atributo	Indica la existencia de competidores del mismo producto en la zona
continua comprando	Atributo	Indica la valoración de si el cliente continuaría comprando
define	Relación	la regla define el movimiento que realiza el cliente
funcionamiento	Atributo	Indica la valoración del cliente respecto el funcionamiento del servicio
gasto	Atributo	Indica la valoración entre el gasto y el producto por parte del cliente
identifica	Relación	el código de cliente identifica cada cliente
identifica	Relación	El código regla identifica a una regla particular
movimientos	Atributo	indica el movimiento realizado por el cliente en los últimos 6 meses
realiza	Relación	el cliente realiza movimientos
recomendación	Atributo	Indica la valoración de si el cliente recomendaría el producto
Regla	Concepto	Variables que definen el movimiento de un cliente
satisfacción general	Atributo	Indica la valoración general del cliente sobre el producto
Subconjunto	Relación	la satisfacción general es una variable del conjunto de la regla
Subconjunto	Relación	la recomendación es una variable del conjunto de la regla
Subconjunto	Relación	continua comprando es una variable del conjunto de la regla
Subconjunto	Relación	cambia proveedor es una variable del conjunto de la regla
Subconjunto	Relación	el funcionamiento es una variable del conjunto de la regla
Subconjunto	Relación	el gasto es una variable del conjunto de la regla
Subconjunto	Relación	la competencia es una variable del conjunto de la regla

Tabla 5.4. Tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) a partir del caso de validación: prestación de servicios de internet

5.1.3.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información

Esta técnica está compuesta por cinco pasos, cuyos elementos de entrada son la descripción de los datos del problema de negocio y del problema de explotación de información, presentadas en la sección 5.1.1, y la tabla término-categoría-definición del problema de explotación de información (TCDPEI) obtenida al ejecutar la técnica tabla término-categoría-definición del problema de explotación de información. El producto resultado del paso es la tabla concepto-atributo-relación-valor extendida del problema de explotación de información (CARVEPEI).

Paso 1. *Identificación de conceptos:* a partir de los términos identificados en la tabla término-categoría-definición del problema de explotación de información (TCDPEI), se registran aquellos elementos de categoría “Concepto” en la columna homónima de la tabla concepto-atributo-relación-valor extendida del problema de explotación de información (CARVEPEI).

Se identifican dos conceptos: “Regla” y “Cliente”.

Paso 2. *Identificación de atributos por concepto:* se lleva a cabo una integración de cada término cuya categoría es “Atributo” identificado en la tabla término-categoría-definición del problema de explotación de información (TCDPEI), al concepto previamente identificado en la tabla concepto-atributo-relación-valor extendida del problema de explotación de información (CARVEPEI).

A continuación se presenta una lista de pares ordenados, en donde el primer elemento corresponde al concepto identificado en el paso previo, y el segundo al atributo asociado a dicho concepto:

(Regla ; código regla), (Regla ; satisfacción general), (Regla ; recomendación), (Regla ; continua comprando), (Regla ; cambia proveedor), (Regla ; funcionamiento), (Regla ; gasto), (Regla ; competencia), (Cliente ; código cliente) y (Cliente ; movimientos).

Paso 3. *Identificación de la relación:* se identifica y registra aquel elemento de la término-categoría-definición del problema de explotación de información (TCDPEI) cuya categoría es “Relación” que relaciona al concepto con el atributo, registrándolo en la columna homónima del atributo correspondiente.

A continuación se presenta una lista de ternas ordenadas, cuyos primeros dos elementos son los identificados en el punto anterior, agregándose el nombre de la relación que los vincula:

(Regla ; código regla ; identifica), (Regla ; satisfacción general ; subconjunto), (Regla ; recomendación ; subconjunto), (Regla ; continua comprando ; subconjunto), (Regla ; cambia proveedor ; subconjunto), (Regla ; funcionamiento ; subconjunto), (Regla ; gasto ; subconjunto), (Regla ; competencia ; subconjunto), (Cliente ; código cliente ; identifica) y (Cliente ; movimientos ; realiza).

Paso 4. *Identificación de los valores posibles:* a partir de la descripción de los datos, se definen los posibles valores de cada atributo identificado.

Se presentan una lista de cuaternas cuyos primeros tres elementos son los identificados en el paso previo, agregando como cuarto elemento los valores posibles de cada atributo:

(Regla ; código regla ; identifica ; Numérico), (Regla ; satisfacción general ; subconjunto ; Satisfecho/Insatisfecho), (Regla ; recomendación ; subconjunto ; Satisfecho/Insatisfecho), (Regla ; continua comprando ; subconjunto ; Satisfecho/Insatisfecho), (Regla ; cambia proveedor ; subconjunto ; Satisfecho/Insatisfecho), (Regla ; funcionamiento ; subconjunto ; Satisfecho/Insatisfecho), (Regla ; gasto ; subconjunto ; Satisfecho/Insatisfecho), (Regla ; competencia ; subconjunto ; Alta/Media/Baja), (Cliente ; código cliente ; identifica ; Numérico) y (Cliente ; movimientos ; realiza ; Alta/Baja/Sin movimientos).

Paso 5. Identificación de los atributos de entrada y salida: a partir de la descripción del problema de explotación de información, se definen cuales atributos son valores de entrada y cuales son valores esperados o salida del problema.

Los datos producidos en los pasos precedentes se integran a la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) junto con los elementos identificados como entrada ó salida del problema. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.5 en la cual se observa al conjunto de términos identificados en la técnica anterior, presentados de una forma organizada la cual permite describir la estructura de cada concepto perteneciente al problema de explotación de información.

Concepto	Atributo	Relación	Entrada/Salida	Valor
Regla	código regla	identifica	Salida	Numérico
	satisfacción general	subconjunto		Satisfecho/Insatisfecho
	recomendación	subconjunto		Satisfecho/Insatisfecho
	continua comprando	subconjunto		Satisfecho/Insatisfecho
	cambia proveedor	subconjunto		Satisfecho/Insatisfecho
	funcionamiento	subconjunto		Satisfecho/Insatisfecho
	gasto	subconjunto		Satisfecho/Insatisfecho
	competencia	subconjunto		Alta/Media/Baja
Cliente	código cliente	identifica		Numérico
	movimientos	realiza	Entrada	Alta/Baja/Sin movimientos

Tabla 5.5. Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) a partir del caso de validación: prestación de servicios de internet

5.1.3.2. Paso 5: “Identificación de relaciones entre conceptos del Problema de Explotación de Información”

En este paso se implementan la técnica *tabla concepto-relación del problema de explotación de información* (sección 5.1.3.2.1), orientada a identificar las interacciones entre los conceptos del problema de explotación de información previamente identificados. Las tablas término-categoría-definición del problema de explotación de información (TCDPEI) y Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI), obtenidas a partir del paso 4, ingresan como elemento de entrada del paso actual, y la tabla concepto-relación del problema de explotación de información (CRPEI) es su producto salida.

5.1.3.2.1. Técnica Tabla Concepto-Relación del Problema de Explotación de Información

Esta técnica está compuesta por dos pasos, y genera como resultado de aplicar dichos pasos a las tablas término-categoría-definición del problema de explotación de información (TCDPEI) y Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) obtenidas en la sección 5.1.3.1, la tabla Concepto-Relación del Problema de Explotación de Información (CRPEI).

Paso 1. *Identificación de relaciones entre conceptos:* se registran aquellos términos de categoría “Relación”, identificados en la tabla término-categoría-definición del problema de explotación de información (TCDPEI), que relacionen a dos conceptos identificados, y se describe a dicha relación.

Se identifica un término de categoría “Relación” que vincula dos conceptos: “define” cuya descripción es “La regla define el movimiento que realiza el cliente”.

Paso 2. *Identificación ordenada de conceptos relacionados:* se registran los conceptos integrantes de las relaciones identificadas en el paso anterior, ingresando al concepto generador de la relación en la columna “Concepto” y al concepto asociado en la columna homónima.

Los datos producidos en el paso previo se integran a la tabla Concepto-Relación del Problema de Explotación de Información (CRPEI) junto con los valores conceptos vinculados. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.6 en la cual se observan las relaciones entre los distintos conceptos que componen al problema.

Concepto	Concepto asociado	Relación	Descripción
Regla	Cliente	define	La regla define el movimiento que realiza el cliente

Tabla 5.6. Tabla Concepto-Relación del Problema de Explotación de Información (CRPEI) a partir del caso de validación: prestación de servicios de internet

5.1.3.3. Paso 6: “Conceptualización del Problema de Explotación de Información”

En este paso se implementan la técnica *Red Semántica del Problema de Explotación de Información* (sección 5.1.3.3.1), cuyo objetivo es conceptualizar la información del problema de explotación de información, desde un enfoque orientado a las relaciones de sus componentes. Las tablas Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) y la Concepto-Relación del Problema de Explotación de Información (CRPEI), obtenidas en los pasos 4 y 5, ingresan como elemento de entrada del paso actual, y la Red Semántica del Problema de Explotación de Información (RSPEI) es su producto salida.

5.1.3.3.1. Técnica Red Semántica del Problema de Explotación de Información

Esta técnica está compuesta por tres pasos, y genera como resultado de aplicar dichos pasos a las tablas Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) y la Concepto-Relación del Problema de Explotación de Información (CRPEI) obtenidas en la sección 5.1.3.1.2 y 5.1.3.2.1 respectivamente, la Red Semántica del Problema de Explotación de Información (RSPEI).

Paso 1. Representación de los conceptos: se representa a partir de la notación definida en la sección 4.2, la estructura de los conceptos identificados en la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI).

Como resultado de aplicar el primer paso se obtiene la figura 5.2.

Paso 2. Representación de las interacciones entre conceptos: se representan las relaciones entre conceptos identificadas en la tabla Concepto-Relación del Problema de Explotación de Información (CRPEI), mediante la anotación provista en la sección 4.2.

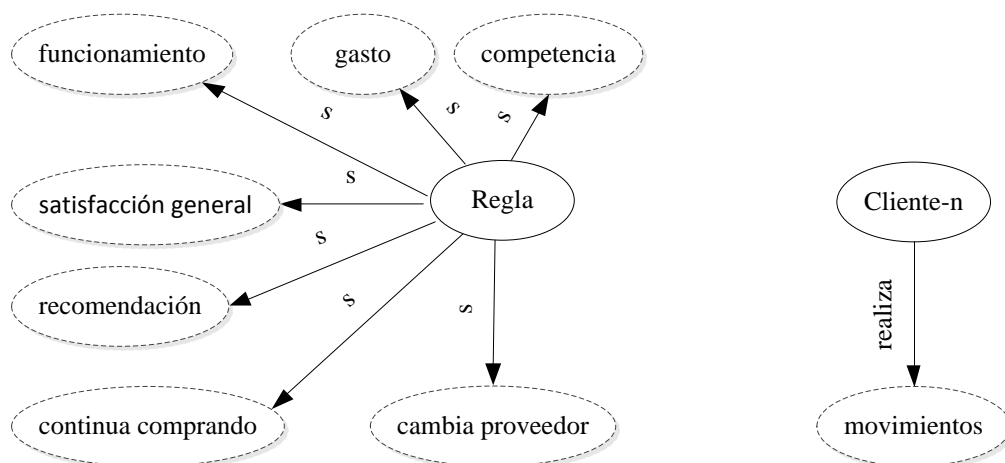


Figura 5.2. Paso 1: Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: prestación de servicios de internet

Como resultado de integrar el segundo paso al paso previo se obtiene la figura 5.3.

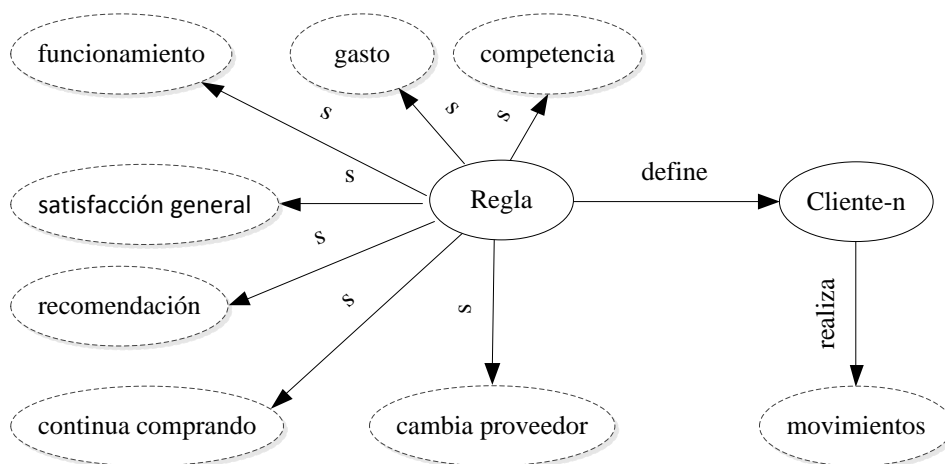


Figura 5.3. Paso 2: Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: prestación de servicios de internet

Paso 3. Representación de los elementos de entrada y salida: se representan los elementos de entrada y salida identificados en la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI), mediante la anotación propuesta en la sección 4.2.2.

El producto obtenido de integrar el tercer paso a los pasos previos, aplicando dichos pasos al presente caso de estudio, se ilustra en la figura 5.4 en la cual se observa al problema de explotación de información descrito a través de las notaciones definidas, desde una perspectiva orientada a las relaciones entre componentes.

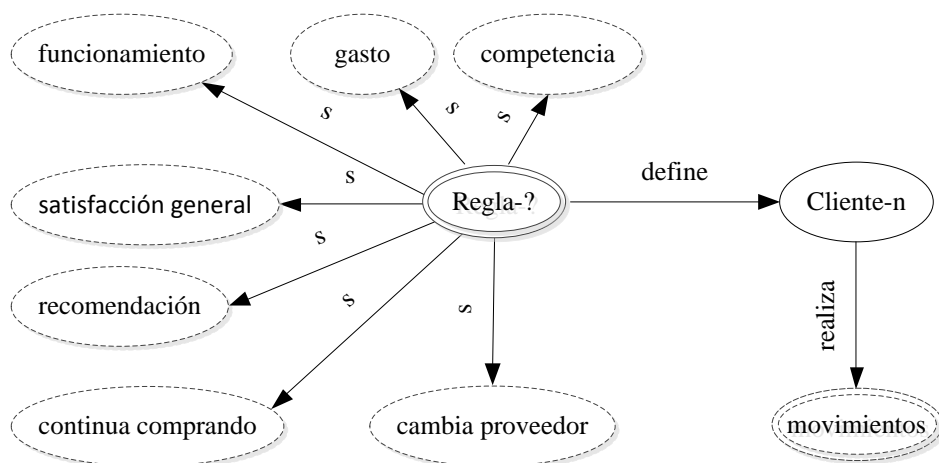


Figura 5.4. Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: prestación de servicios de internet

5.1.4. Aplicación de la Técnica Utilizada en la Fase de Determinación del Proceso de Explotación de Información

En esta sección se aplican al presente caso de validación la etapa “*derivación del proceso de explotación de información*” (sección 5.1.4.1), cuyo resultado esperado es el resultado final del proceso propuesto.

5.1.4.1. Paso 7: “Derivación del Proceso de Explotación de Información”

En este paso se implementan la técnica *algoritmo de derivación del proceso de explotación de información* (sección 5.1.4.1.1), cuyo objetivo es identificar el proceso de explotación de información a aplicar en base a las representaciones intermedias obtenidas. La Red Semántica del Modelo de Negocio (RSMN), generada en el paso 3, y la Red Semántica del Problema de Explotación de Información (RSPEI), generado en el paso 6, son los elementos de entrada del paso, siendo el elemento de salida el proceso de explotación de información a aplicar.

5.1.4.1.1. Técnica Algoritmo de Derivación del Proceso de Explotación de Información

Esta técnica está compuesta por un único paso, el cual está conformado por ocho subpasos condicionales, y genera a partir la Red Semántica del Modelo de Negocio (sección 5.1.2.3) y la Red Semántica del Problema de Explotación de Información (sección 5.1.3.3) aplicadas a la técnica propuesta, el proceso de explotación de información a aplicar.

Paso 1. Ejecución del Algoritmo de derivación: se identifica en las redes semánticas que conforman el conjunto de elementos de entrada de la técnica, los patrones definidos en cada subpaso del Algoritmo de Derivación del Proceso de Explotación de Información (ADPEI).

La secuencia de pasos obtenida de la ejecución del presente caso de estudio en el Algoritmo de Derivación del Proceso de Explotación de Información (ADPEI) se presenta a continuación, junto con el proceso de explotación de información identificado como resultado. Para la comprensión de los elementos detectados en cada subpaso, se aplica la siguiente notación: líneas punteadas para arcos o nodos identificados como respuesta en cada subpaso, y subrayado para texto.

Subpaso 7.1: Se identifica un nodo variable en la Red Semántica del Problema de Explotación de Información (figura 5.5).

✓ Resultado obtenido: Ir a Subpaso 7.2

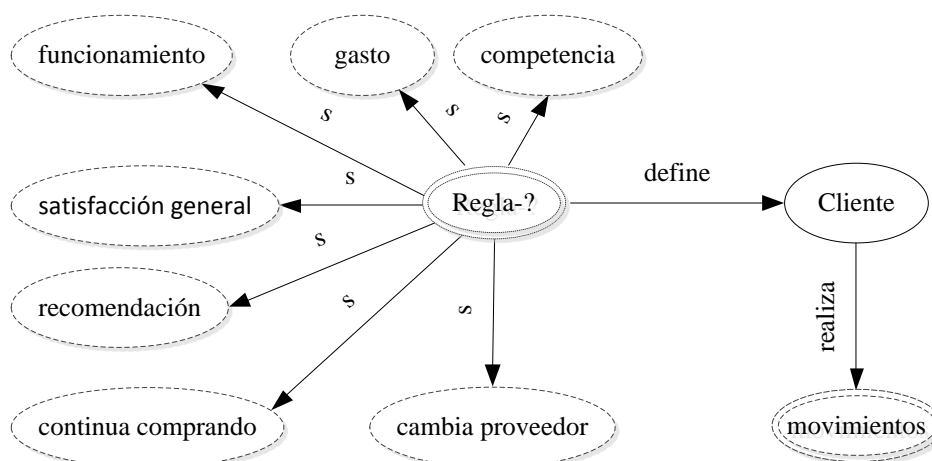


Figura 5.5. Identificación de subpaso 7.1 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: prestación de servicios de internet

Subpaso 7.2: No se identifica un nodo variable en la Red Semántica del Problema de Explotación de Información etiquetado con las palabras: pertenece, integra, es, es miembro de, u otras palabras de similar significado.

✓ Resultado obtenido: Ir a Subpaso 7.3

Subpaso 7.3: No se identifica un nodo variable en la Red Semántica del Problema de Explotación de Información etiquetado con las palabras: elementos claves, factores destacados, elementos relevantes u otras palabras de similar significado.

✓ Resultado obtenido: Ir a Subpaso 7.4

Subpaso 7.4: Se identifica un nodo variable, en la Red Semántica del Problema de Explotación de Información, etiquetado con la palabra “regla” (figura 5.6).

✓ Resultado obtenido: **Aplicar Proceso de Descubrimiento de Reglas de Comportamiento**

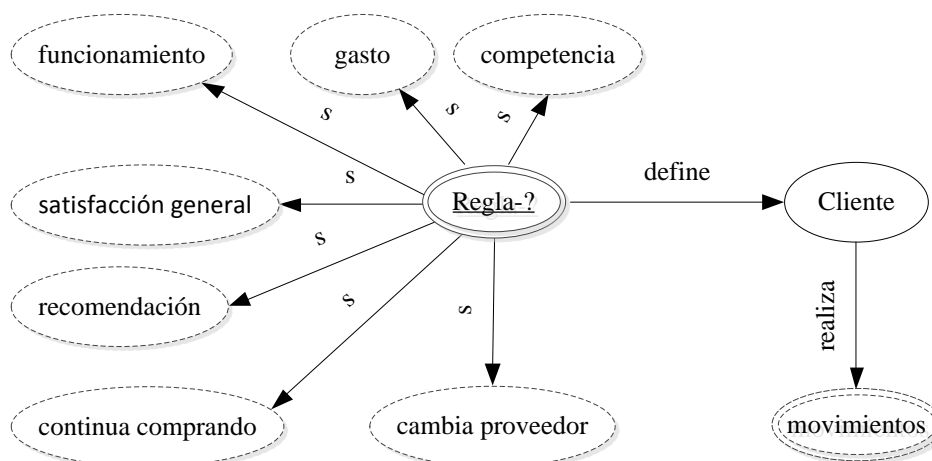


Figura 5.6. Identificación de subpaso 7.4 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: prestación de servicios de internet

Retornando como respuesta del paso y del *proceso de derivación de modelos* aplicado al caso de validación: *prestación de servicios de internet*, el texto:

“Proceso de Descubrimiento de Reglas de Comportamiento”.

5.2. CASO DE VALIDACIÓN: INDUSTRIA AUTOMOTRIZ

En esta sección se analiza el segundo caso de validación correspondiente a una empresa perteneciente a la industria automotriz. En la sección 5.2.1 se describe al modelo de negocio, sus datos y el problema de explotación de información a analizar. En la sección 5.2.2 se aplican las etapas y técnicas utilizadas para el desarrollo de las tareas correspondientes a la fase de *análisis del dominio del negocio*. En la sección 5.2.3 se aplican las etapas y técnicas utilizadas para el desarrollo de las tareas correspondientes a la fase de *análisis del problema de explotación de información*. En la sección 5.2.4 se aplican la etapa y técnica utilizada para el desarrollo de las tareas correspondientes a la fase de *determinación del proceso de explotación de información* obteniendo como resultado el proceso de explotación de información a aplicar en base al modelo del negocio y el problema analizado.

5.2.1. Descripción

En esta sección se presenta la información necesaria para comprender el modelo de negocio y el problema de explotación de información a analizar. El presente caso se obtiene de [Flores, 2009], y se presentan las características relevantes del mismo:

“...Descripción del Negocio y Problema de Negocio:

En todo el circuito de distribución de automóviles existe la posibilidad y el riesgo de provocar daños a los automóviles.

A lo largo de la cadena logística de este circuito se pasa por diversos puntos donde circulan los autos desde que salen de producción hasta que llegan a su destino...

En cada uno de los puntos de control (plantas, puertos, playas, etc.) para la realización de las operaciones de carga o descarga, ingreso o salida, los operadores logísticos controlan que es lo que sucede con los autos en el cambio de responsabilidad cuando se traspasa una unidad de una empresa a otra, debido a que los daños y/o averías que sufren los mismos deben ser imputados (asignados) a la/s empresa/s que lo/s produjeron.

Los automóviles son identificados mediante un valor entero, registrando además su modelo.

Los daños son registrados según su tipo, su categoría (la cual indica el grado de severidad del mismo), además de tener una relación directamente proporcional a la magnitud del costo que el daño genera y la fecha en que se produce.

Es de vital importancia para la empresa identificar aquellas características que inciden en mayor forma a la generación de averías, en base a su gravedad o seriedad, principalmente en aquellas con mayor gravedad que generan un mayor costo de reparación. De esta forma, se busca disminuir el gasto en reparaciones.

Problema de explotación de información

Determinar en base a los factores de traslado y del tipo de producto trasladado, las características que tienen mayor incidencia en las averías del vehículo, según su gravedad.

Descripción de los datos

Para ello la empresa cuenta con la información vinculada a los temas de siniestralidad de vehículos, almacenada en forma distribuida entre las distintas empresas responsables en cada punto de control de cada cadena logística, los datos que se registran son:

- *Identificador de auto (entero): Código único que identifica los automóviles,*
- *Fecha (aaaa/mm/dd): fecha de operación en la que se detecta el daño.*
- *Modelo (entero): identificador del modelo del automóvil,*
- *Lugar (entero): Lugar donde se produce la falla,*
- *Autoparte (entero): Parte del automóvil donde se produce la avería,*
- *Id avería (entero): identificador de la avería,*
- *Tipo avería (entero): Tipo de avería producida en cada unidad,*
- *Gravedad (entero): Severidad o gravedad de la avería producida, y*
- *Transporte (entero): Transporte en el que se produjo la avería.*
- *Empresa (entero): Empresa dueña del transporte...”*

5.2.2. Aplicación de las Técnicas Utilizadas en la Fase de Análisis del Dominio del Negocio

En esta sección se aplican a este caso de validación las etapas y técnicas utilizadas para el desarrollo de las tareas correspondientes a la fase de análisis del dominio del negocio. Las etapas que componen la fase son: *identificación de los elementos y estructura del dominio* (sección 5.2.2.1),

identificación de relaciones entre conceptos del dominio (sección 5.2.2.2) y *conceptualización del dominio* (sección 5.2.2.3).

5.2.2.1. Paso 1: “Identificación de los elementos y estructura del dominio”

En este paso se implementan dos técnicas: la *tabla término-categoría-definición del dominio* (sección 5.2.2.1.1), orientada a identificar los elementos relevantes del modelo de negocio y la *tabla concepto-atributo-relación-valor del dominio* (sección 5.2.2.1.2), cuyo objetivo es definir la estructura de los conceptos relevantes que integran al negocio. La descripción del dominio y los datos del negocio ingresan como elemento de entrada del paso, y las tablas término-categoría-definición del dominio (TCDD) y tabla concepto-atributo-relación-valor del dominio (CARVD), son los productos salida del paso.

5.2.2.1.1. Técnica Tabla Término-Categoría-Definición del Dominio

Esta técnica está compuesta por tres pasos, y genera como resultado de aplicar dichos pasos a la descripción del negocio provista en la sección 5.2.1, la tabla término-categoría-definición del dominio (TCDD).

Paso 1. *Identificación de términos relevantes:* se lleva a cabo el análisis de la descripción del negocio a los efectos de identificar y registrar alfabéticamente los elementos que permiten comprender el negocio.

A continuación se presenta el listado de términos identificados, ordenado alfabéticamente: "Automóvil", "autoparte", "Avería", "Código automóvil", "código avería", "Código empresa", "detectada en", "Empresa", "es", "es", "fecha", "gravedad", "identifica", "identifica", "identifica", "modelo", "pertenece", "produce", "producida en", "punto de control", "se produce", "tiene", "tiene", "tipo" y "tipo de transporte".

Paso 2. *Clasificación de términos por categoría:* a partir de los términos identificados en el paso anterior, se los clasifica de acuerdo al tipo de elemento que representa (sección 4.3.2.1).

A continuación se presenta el listado de términos identificados en el paso anterior, junto con la categoría asignada a cada uno de ellos, representando dicha asignación mediante pares ordenados, donde el primer elemento es el nombre del término y el segundo la categoría designada:

(Automóvil , Concepto), (autoparte , Atributo), (Avería , Concepto), (Código automóvil , Atributo), (código avería , Atributo), (Código empresa , Atributo), (detectada en , Relación), (Empresa , Concepto), (es , Relación), (es , Atributo), (fecha , Atributo), (gravedad , Atributo), (identifica , Relación), (identifica , Relación), (identifica , Relación), (modelo , Atributo), (pertenece , Relación), (produce , Relación), (producida en , Relación), (punto de control , Atributo), (se produce , Relación), (tiene , Relación), (tiene , Relación), (tipo , Atributo) y (tipo de transporte , Atributo).

Paso 3. Describir el término: se lleva a cabo una descripción de cada término identificado, en base al dominio del negocio que integra.

Los datos producidos en los pasos precedentes se integran a la tabla Término-Categoría-Definición del Dominio (TCDD) junto con la descripción de cada concepto. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.7 en la cual se observa el conjunto de términos identificados, su categoría y la descripción que los define de acuerdo al negocio analizado.

Término	Categoría	Definición
Automóvil	Concepto	Producto transportado
autoparte	Atributo	parte del automóvil en donde se registró la avería
Avería	Concepto	daño producido en una parte del automóvil
Código automóvil	Atributo	Clave única que representa cada automóvil
código avería	Atributo	Identificador único por avería
Código empresa	Atributo	Clave única que representa cada empresa
detectada en	Relación	La avería ocurre en un punto de control
Empresa	Concepto	Entidad responsable de la avería producida
es	Relación	la empresa es de un tipo de transporte
es	Atributo	la avería es de un tipo específico
fecha	Atributo	día en que se produce la avería
gravedad	Atributo	indicador del grado de severidad de la avería producida
identifica	Relación	El automóvil se identifica mediante el código de automóvil
identifica	Relación	La empresa se identifica mediante el código de empresa
identifica	Relación	La avería se identifica mediante el código de avería
modelo	Atributo	Modelo del automóvil
pertenece	Relación	la avería se registra en un automóvil
produce	Relación	la empresa produce la avería
producida en	Relación	La avería ocurre en una autoparte
punto de control	Atributo	Lugar donde se ocasiona la avería
se produce	Relación	la avería se produce en una fecha
tiene	Relación	El automóvil tiene un modelo
tiene	Relación	la avería tiene una gravedad
tipo	Atributo	Tipo de avería producida
tipo de transporte	Atributo	tipo de transporte de la empresa en el que se produce la avería

Tabla 5.7. Tabla Término-Categoría-Definición del Dominio (TCDD) a partir del caso de validación: industria automotriz

5.2.2.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio

Esta técnica está compuesta por cuatro pasos, cuyos elementos de entrada son la descripción de los datos del negocio provista en la sección 5.2.1, y la tabla término-categoría-definición del dominio (TCDD) obtenida de ejecutar la técnica tabla término-categoría-definición del dominio. El producto resultado del paso es la tabla concepto-atributo-relación-valor del dominio (CARVD).

Paso 1. *Identificación de conceptos:* a partir de los términos identificados en la tabla término-categoría-definición del dominio (TCDD), se registran aquellos elementos de categoría “Concepto” en la columna homónima de la tabla concepto-atributo-relación-valor del dominio (CARVD).

Se identifican tres elementos de categoría “Concepto”: Automóvil, Avería y Empresa.

Paso 2. *Identificación de atributos por concepto:* se lleva a cabo una integración de cada término cuya categoría es “Atributo” identificado en la tabla término-categoría-definición del dominio (TCDD), al concepto previamente identificado en la tabla concepto-atributo-relación-valor del dominio (CARVD).

A continuación se presenta una lista de pares ordenados, en donde el primer elemento corresponde al concepto identificado en el paso previo, y el segundo al atributo asociado al concepto:

(Automóvil , Código automóvil), (Automóvil , modelo), (Avería , Código avería), (Avería , autoparte), (Avería , tipo), (Avería , fecha), (Avería , gravedad), (Avería , punto de control), (Empresa , Código empresa) y (Empresa , tipo de transporte).

Paso 3. *Identificación de la relación:* se identifica y registra aquel elemento de la tabla término-categoría-definición del dominio (TCDD) cuya categoría es “Relación” que relaciona al concepto con el atributo, registrándolo en la columna homónima del atributo correspondiente.

A continuación se presenta una lista de ternas ordenadas, cuyos primeros dos elementos son los identificados en el punto anterior, agregándose el nombre de la relación que los vincula:

(Automóvil , Código automóvil , identifica), (Automóvil , modelo , tiene), (Avería , Código avería , identifica), (Avería , autoparte , producida en), (Avería , tipo , es), (Avería , fecha , se produce),

(Avería , gravedad , tiene), (Avería , punto de control , detectada en), (Empresa , Código empresa , identifica) y (Empresa , tipo de transporte , es).

Paso 4. Identificación de los valores posibles: a partir de la descripción de los datos, se definen los posibles valores de cada atributo identificado.

Los datos producidos en los pasos precedentes se integran a la tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) junto con los valores posibles de cada atributo. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.8 en la cual se observa al conjunto de términos identificados en la técnica anterior, presentados de una forma organizada la cual permite describir la estructura de cada concepto perteneciente al dominio del negocio.

Concepto	Atributo	Relación	Valor
Cliente	código cliente	identifica	Numérico
	satisfacción general	presenta	Satisfecho/Insatisfecho
	recomendación	tiene	Satisfecho/Insatisfecho
	continúa comprando	posee	Satisfecho/Insatisfecho
	cambia proveedor	posee	Satisfecho/Insatisfecho
	funcionamiento	describe	Satisfecho/Insatisfecho
	gasto	describe	Satisfecho/Insatisfecho
	competencia	hay	Alta/Media/Baja
	movimientos	realiza	Alta/Baja/Sin movimientos

Tabla 5.8. Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) a partir del caso de validación: industria automotriz

5.2.2.2. Paso 2: “Identificación de Relaciones entre Conceptos del Dominio”

En este paso se implementan la técnica *tabla concepto-relación del dominio* (sección 5.2.2.2.1), orientada a identificar las interacciones entre los conceptos del dominio de negocio previamente identificados. Las tablas término-categoría-definición del dominio (TCDD) y concepto-atributo-relación-valor del dominio (CARVD), obtenidas a partir del paso 1, ingresan como elemento de entrada del paso actual, y la tabla concepto-relación del dominio (CRD) es su producto salida.

5.2.2.2.1. Técnica Tabla Concepto-Relación del Dominio

Esta técnica está compuesta por dos pasos, y genera como resultado de aplicar dichos pasos a las tablas término-categoría-definición del dominio (TCDD) y concepto-atributo-relación-valor del dominio (CARVD) obtenidas en la sección 5.2.2.1, la tabla concepto-relación del dominio (CRD).

Paso 1. Identificación de relaciones entre conceptos: se registran aquellos términos de categoría “Relación”, identificados en la tabla término-categoría-definición del dominio (TCDD), que relacionen a dos conceptos identificados, y se describe a dicha relación.

Se identifican dos términos que vinculan conceptos: “pertenece” y “produce”, cuyas descripciones son “la avería registrada pertenece a un automóvil” y “la empresa de transporte produce la avería” respectivamente.

Paso 2. Identificación ordenada de conceptos relacionados: se registran los conceptos integrantes de las relaciones identificadas en el paso anterior, ingresando al concepto generador de la relación en la columna “Concepto” y al concepto asociado en la columna homónima.

El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.9 en la cual se observan los conceptos relacionados, su relación y una descripción de la misma en base al dominio del negocio.

Conceptos	Conceptos asociados	Relación	Descripción
avería	automóvil	pertenece	la avería registrada pertenece a un automóvil
empresa	avería	produce	la empresa de transporte produce la avería

Tabla 5.9. Tabla Concepto-Relación del Dominio (CRD) a partir del caso de validación: industria automotriz

5.2.2.3. Paso 3: “Conceptualización Del Dominio”

En este paso se implementan la técnica *red semántica del modelo de negocio* (sección 5.2.2.3.1), cuyo objetivo es conceptualizar la información del dominio del negocio, desde un enfoque orientado a las relaciones de sus componentes. Las tablas concepto-atributo-relación-valor del dominio (CARVD) y la tabla concepto-relación del dominio (CRD), obtenidas en los pasos 1 y 2 respectivamente, ingresan como elemento de entrada del paso actual, y la red semántica del modelo de negocio (RSMN) es su producto salida.

5.2.2.3.1. Técnica Red Semántica del Modelo de Negocio

Esta técnica está compuesta por dos pasos, y genera como resultado de aplicar dichos pasos a las tablas concepto-atributo-relación-valor del dominio (CARVD) y concepto-relación del dominio (CRD) obtenidas en la sección 5.2.2.1.2 y 5.2.2.2.1 respectivamente, la red semántica del modelo de negocio (RSMN).

Paso 1. Representación de los conceptos: se representa a partir de la notación definida en la sección 4.2, la estructura de los conceptos identificados en la tabla concepto-atributo-relación-valor del dominio (CARVD).

El producto obtenido de aplicar dicho paso al presente caso de estudio, se ilustra en la figura 5.7 en la cual se observa al modelo de negocio descrito a través de las notaciones definidas, desde una perspectiva orientada a las relaciones entre componentes.

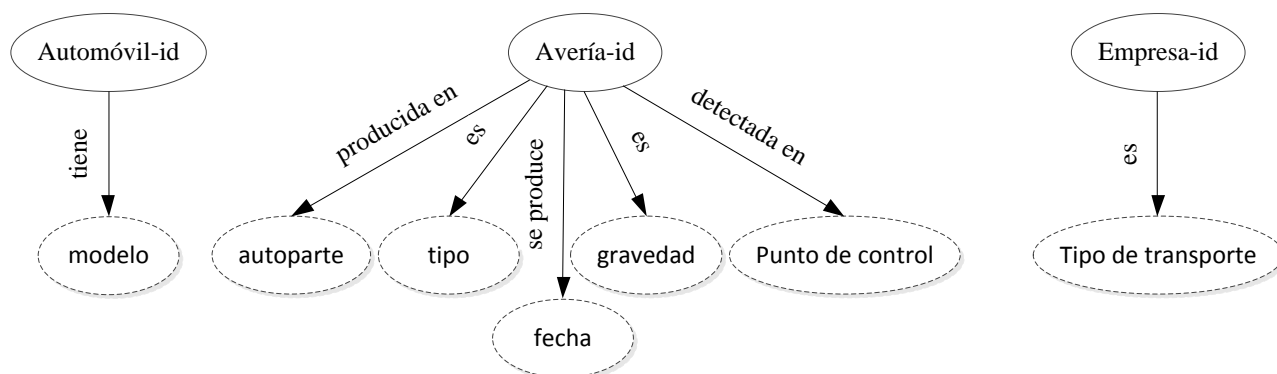


Figura 5.7. Paso 1: Red Semántica del Modelo de Negocio (RSMN) a partir del caso de validación: industria automotriz

Paso 2. Representación de las interacciones entre conceptos: se representan las relaciones entre conceptos identificadas en la tabla concepto-relación del dominio (CRD), mediante la anotación provista en la sección 4.2.

El producto obtenido de integrar el segundo paso, al paso previo aplicando dichos pasos al presente caso de estudio, se ilustra en la figura 5.8 en la cual se observa al modelo del negocio descrito a través de las notaciones definidas, desde una perspectiva orientada a las relaciones entre componentes.

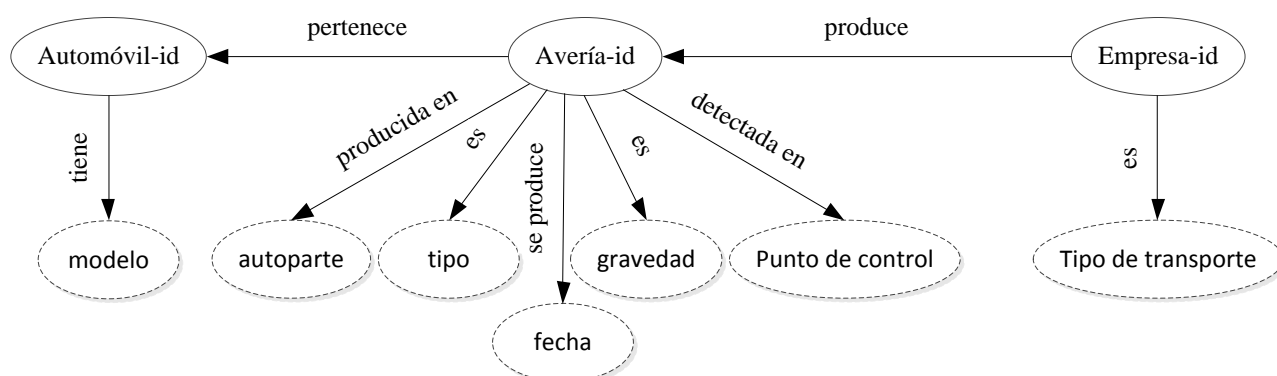


Figura 5.8. Red Semántica del Modelo de Negocio (RSMN) a partir del caso de validación: industria automotriz

5.2.3. Aplicación de las Técnicas Utilizadas en la Fase de Análisis del Problema de Explotación de Información

En esta sección se aplican al presente caso de validación las etapas y técnicas utilizadas para el desarrollo de las tareas correspondientes a la fase de análisis del problema de explotación de información. Las etapas que componen la fase son: *identificación de los elementos y estructura del problema de explotación de información* (sección 5.2.3.1), *identificación de relaciones entre conceptos del problema de explotación de información* (sección 5.2.3.2) y *conceptualización del problema de explotación de información* (sección 5.2.3.3).

5.2.3.1. Paso 4: “Identificación de los elementos y estructura del Problema de Explotación de Información”

En este paso se implementan dos técnicas: la *tabla término-categoría-definición del problema de explotación de información* (sección 5.2.3.1.1), orientada a identificar los elementos relevantes del problema de explotación de información y la *tabla concepto-atributo-relación-valor del problema de explotación de información* (sección 5.2.3.1.2), cuyo objetivo es definir la estructura de los conceptos relevantes que integran al problema. La descripción y los datos del problema ingresan como elemento de entrada del paso, y las tablas término-categoría-definición del problema de explotación de información (TCDPEI) y tabla concepto-atributo-relación-valor extendida del problema de explotación de información (CARVEPEI), son los productos salida del paso.

5.2.3.1.1. Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información

Esta técnica está compuesta por tres pasos, y genera como resultado de aplicar dichos pasos a la descripción del problema de explotación de información provista en la sección 5.2.1, la tabla término-categoría-definición del problema de explotación de información (TCDPEI).

Paso 1. *Identificación de términos relevantes:* se lleva a cabo el análisis de la descripción del problema de explotación de información a los efectos de identificar y registrar alfabéticamente los elementos que permiten comprender al mismo.

A continuación se presenta el listado de términos identificados, ordenado alfabéticamente:

"características relevantes", "autoparte", "Avería", "código avería", "código características relevantes", "definen", "gravedad", "identifica", "identifica", "modelo", "punto de control", "según", "Subconjunto", "Subconjunto", "Subconjunto", "Subconjunto", "Subconjunto", "tipo de avería" y "tipo de transporte".

Paso 2. Clasificación de términos por categoría: a partir de los términos identificados en el paso anterior, se los clasifica de acuerdo al tipo de elemento que representa (sección 4.3.2.1).

A continuación se presenta el listado de términos identificados en el paso anterior, junto con la categoría asignada a cada uno de ellos, representando dicha asignación mediante pares ordenados, donde el primer elemento es el nombre del término y el segundo la categoría designada:

(características relevantes , Concepto), (autoparte , Atributo), (Avería , Concepto), (código avería , Atributo), (código características relevantes , Atributo), (definen , Relación), (gravedad , Concepto), (identifica , Relación), (identifica , Relación), (modelo , Atributo), (punto de control , Atributo), (según , Relación), (Subconjunto , Relación), (Subconjunto , Relación), (Subconjunto , Relación), (Subconjunto , Relación), (Subconjunto , Relación), (tipo de avería , Atributo) y (tipo de transporte , Atributo).

Paso 3. Describir el término: se lleva a cabo una descripción de cada término identificado, en base al problema de explotación de información que integra.

Los datos producidos en los pasos precedentes se integran a la tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) junto con la descripción de cada concepto. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.10 en la cual se observa el conjunto de términos identificados, su categoría y la descripción que los define en base al problema de explotación de información analizado.

5.2.3.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información

Esta técnica está compuesta por cinco pasos, cuyos elementos de entrada son la descripción de los datos del problema de negocio y del problema de explotación de información, presentadas en la sección 5.2.1, y la tabla término-categoría-definición del problema de explotación de información (TCDPEI) obtenida al ejecutar la técnica tabla término-categoría-definición del problema de explotación de información.

Término	Categoría	Definición
cambia proveedor	Atributo	Indica la valoración de si el cliente cambiaría de proveedor del servicio
Cliente	Concepto	Persona que contrata alguno de los servicios de la empresa
código cliente	Atributo	Identificador de cada cliente
Código regla	Atributo	Identificación univoca de la regla
competencia	Atributo	Indica la existencia de competidores del mismo producto en la zona
continua comprando	Atributo	Indica la valoración de si el cliente continuaría comprando
define	Relación	la regla define el movimiento que realiza el cliente
funcionamiento	Atributo	Indica la valoración del cliente respecto el funcionamiento del servicio
gasto	Atributo	Indica la valoración entre el gasto y el producto por parte del cliente
identifica	Relación	el código de cliente identifica cada cliente
identifica	Relación	El código regla identifica a una regla particular
movimientos	Atributo	indica el movimiento realizado por el cliente en los últimos 6 meses
realiza	Relación	el cliente realiza movimientos
recomendación	Atributo	Indica la valoración de si el cliente recomendaría el producto
Regla	Concepto	Variables que definen el movimiento de un cliente
satisfacción general	Atributo	Indica la valoración general del cliente sobre el producto
Subconjunto	Relación	la satisfacción general es una variable del conjunto de la regla
Subconjunto	Relación	la recomendación es una variable del conjunto de la regla
Subconjunto	Relación	continua comprando es una variable del conjunto de la regla
Subconjunto	Relación	cambia proveedor es una variable del conjunto de la regla
Subconjunto	Relación	el funcionamiento es una variable del conjunto de la regla
Subconjunto	Relación	el gasto es una variable del conjunto de la regla
Subconjunto	Relación	la competencia es una variable del conjunto de la regla

Tabla 5.10. Tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) a partir del caso de validación: industria automotriz

El producto resultado del paso es la tabla concepto-atributo-relación-valor extendida del problema de explotación de información (CARVEPEI).

Paso 1. Identificación de conceptos: a partir de los términos identificados en la tabla término-categoría-definición del problema de explotación de información (TCDPEI), se registran aquellos elementos de categoría “Concepto” en la columna homónima de la tabla concepto-atributo-relación-valor extendida del problema de explotación de información (CARVEPEI).

Se identifican dos conceptos: “Avería” y “Características relevantes”.

Paso 2. Identificación de atributos por concepto: se lleva a cabo una integración de cada término cuya categoría es “Atributo” identificado en la tabla término-categoría-definición del problema de explotación de información (TCDPEI), al concepto previamente identificado en la tabla concepto-atributo-relación-valor extendida del problema de explotación de información (CARVEPEI).

A continuación se presenta una lista de pares ordenados, en donde el primer elemento corresponde al concepto identificado en el paso previo, y el segundo al atributo asociado a dicho concepto:

(Características relevantes , código características relevantes), (Características relevantes , modelo), (Características relevantes , punto de control), (Características relevantes , autoparte), (Características relevantes , tipo de transporte), (Características relevantes , tipo de avería), (Avería , código avería) y (Avería , gravedad).

Paso 3. *Identificación de la relación:* se identifica y registra aquel elemento de la término-categoría-definición del problema de explotación de información (TCDPEI) cuya categoría es “Relación” que relaciona al concepto con el atributo, registrándolo en la columna homónima del atributo correspondiente.

A continuación se presenta una lista de ternas ordenadas, cuyos primeros dos elementos son los identificados en el punto anterior, agregándose el nombre de la relación que los vincula:

(Características relevantes , código características relevantes , identifica), (Características relevantes , modelo , subconjunto), (Características relevantes , punto de control , subconjunto), (Características relevantes , autoparte , subconjunto), (Características relevantes , tipo de transporte , subconjunto), (Características relevantes , tipo de avería , subconjunto), (Avería , código avería , identifica) y (Avería , gravedad , según).

Paso 4. *Identificación de los valores posibles:* a partir de la descripción de los datos, se definen los posibles valores de cada atributo identificado.

Se presentan una lista de cuaternas cuyos primeros tres elementos son los identificados en el paso previo, agregando como cuarto elemento los valores posibles de cada atributo:

(Características relevantes , código características relevantes , identifica , Numérico), (Características relevantes , modelo , subconjunto , Numérico), (Características relevantes , punto de control , subconjunto , Numérico), (Características relevantes , autoparte , subconjunto , Numérico), (Características relevantes , tipo de transporte , subconjunto , Numérico), (Características relevantes , tipo de avería , subconjunto , Numérico), (Avería , código avería , identifica , Numérico) y (Avería , gravedad , según , Numérico).

Paso 5. *Identificación de los atributos de entrada y salida:* a partir de la descripción del problema de explotación de información, se definen cuales atributos son valores de entrada y cuales son valores esperados o salida del problema.

Los datos producidos en los pasos precedentes se integran a la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) junto con los elementos identificados como entrada ó salida del problema. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.11 en la cual se observa al conjunto de términos identificados en la técnica anterior, presentados de una forma organizada la cual permite describir la estructura de cada concepto perteneciente al problema de explotación de información.

Concepto	Atributo	Relación	Entrada/Salida	Valor
Características relevantes	código características relevantes	identifica	Salida	Numérico
	modelo	subconjunto		Numérico
	punto de control	subconjunto		Numérico
	autoparte	subconjunto		Numérico
	tipo de transporte	subconjunto		Numérico
	tipo de avería	subconjunto		Numérico
Avería	código avería	identifica		Numérico
	gravedad	según	Entrada	Numérico

Tabla 5.11. Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) a partir del caso de validación: industria automotriz

5.2.3.2. Paso 5: “Identificación de relaciones entre conceptos del Problema de Explotación de Información”

En este paso se implementan la técnica *tabla concepto-relación del problema de explotación de información* (sección 5.2.3.2.1), orientada a identificar las interacciones entre los conceptos del problema de explotación de información previamente identificados. Las tablas término-categoría-definición del problema de explotación de información (TCDPEI) y Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI), obtenidas a partir del paso 4, ingresan como elemento de entrada del paso actual, y la tabla concepto-relación del problema de explotación de información (CRPEI) es su producto salida.

5.2.3.2.1. Técnica Tabla Concepto-Relación del Problema de Explotación de Información

Esta técnica está compuesta por dos pasos, y genera como resultado de aplicar dichos pasos a las tablas término-categoría-definición del problema de explotación de información (TCDPEI) y Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) obtenidas en la sección 5.2.3.1, la tabla Concepto-Relación del Problema de Explotación de Información (CRPEI).

Paso 1. Identificación de relaciones entre conceptos: se registran aquellos términos de categoría “Relación”, identificados en la tabla término-categoría-definición del problema de explotación de información (TCDPEI), que relacionen a dos conceptos identificados, y se describe a dicha relación.

Se identifica un término de categoría “Relación” que vincula dos conceptos: “definen” cuya descripción es “las características relevantes definen la avería”.

Paso 2. Identificación ordenada de conceptos relacionados: se registran los conceptos integrantes de las relaciones identificadas en el paso anterior, ingresando al concepto generador de la relación en la columna “Concepto” y al concepto asociado en la columna homónima.

Los datos producidos en el paso previo se integran a la tabla Concepto-Relación del Problema de Explotación de Información (CRPEI) junto con los valores conceptos vinculados. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.12 en la cual se observan las relaciones entre los distintos conceptos que componen al problema.

Concepto	Concepto asociado	Relación	Descripción
Regla	Cliente	define	La regla define el movimiento que realiza el cliente

Tabla 5.12. Tabla Concepto-Relación del Problema de Explotación de Información (CRPEI) a partir del caso de validación: industria automotriz

5.2.3.3. Paso 6: “Conceptualización del Problema de Explotación de Información”

En este paso se implementan la técnica *Red Semántica del Problema de Explotación de Información* (sección 5.2.3.3.1), cuyo objetivo es conceptualizar la información del problema de explotación de información, desde un enfoque orientado a las relaciones de sus componentes. Las tablas Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) y la Concepto-Relación del Problema de Explotación de Información (CRPEI), obtenidas en los pasos 4 y 5, ingresan como elemento de entrada del paso actual, y la Red Semántica del Problema de Explotación de Información (RSPEI) es su producto salida.

5.2.3.3.1. Técnica Red Semántica del Problema de Explotación de Información

Esta técnica está compuesta por tres pasos, y genera como resultado de aplicar dichos pasos a las tablas Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información

(CARVEPEI) y la Concepto-Relación del Problema de Explotación de Información (CRPEI) obtenidas en la sección 5.2.3.1.2 y 5.2.3.2.1 respectivamente, la Red Semántica del Problema de Explotación de Información (RSPEI).

Paso 1. Representación de los conceptos: se representa a partir de la notación definida en la sección 4.2, la estructura de los conceptos identificados en la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI).

Como resultado de aplicar el primer paso se obtiene la figura 5.9.

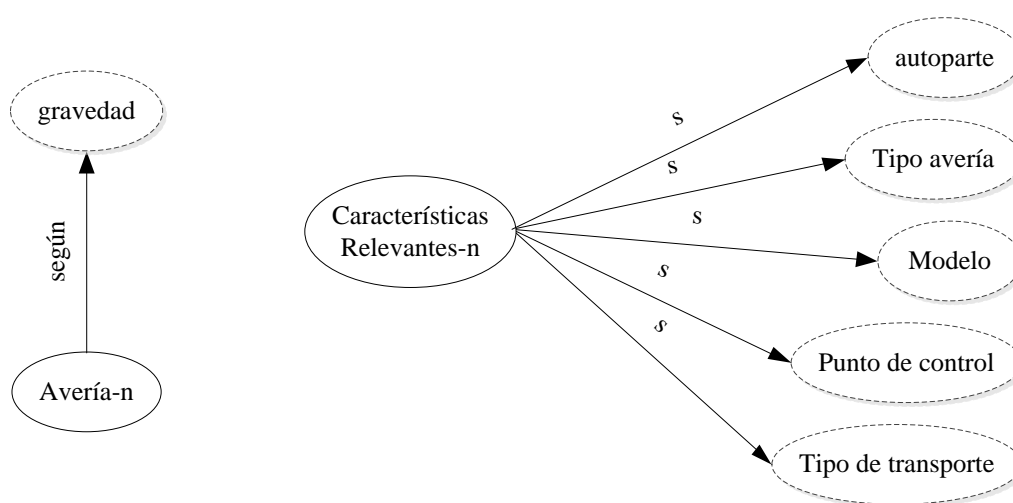


Figura 5.9. Paso 1: Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: industria automotriz

Paso 2. Representación de las interacciones entre conceptos: se representan las relaciones entre conceptos identificadas en la tabla Concepto-Relación del Problema de Explotación de Información (CRPEI), mediante la anotación provista en la sección 4.2.

Como resultado de integrar el segundo paso al paso previo se obtiene la figura 5.10.

Paso 3. Representación de los elementos de entrada y salida: se representan los elementos de entrada y salida identificados en la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI), mediante la anotación propuesta en la sección 4.2.2.

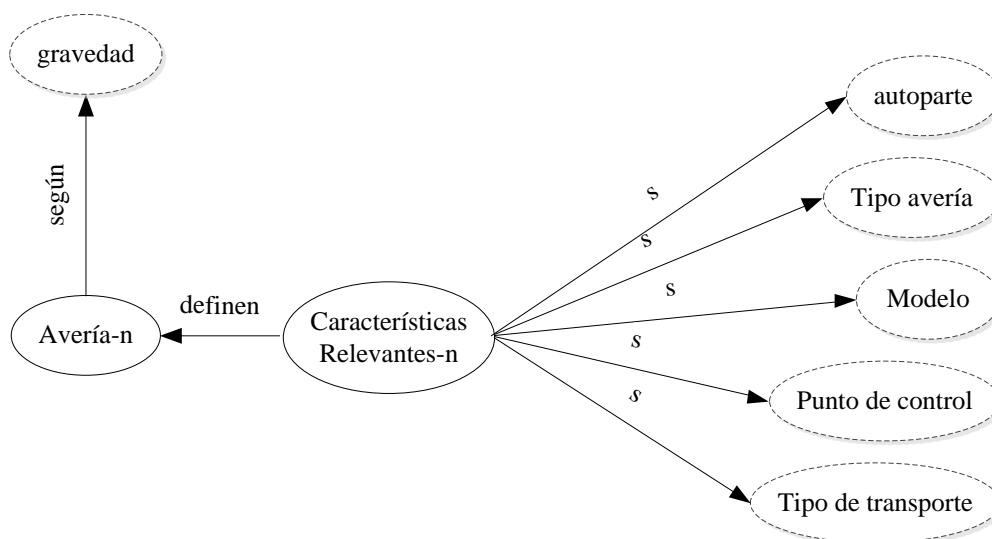


Figura 5.10. Paso 2: Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: industria automotriz

El producto obtenido de integrar el tercer paso a los pasos previos, aplicando dichos pasos al presente caso de estudio, se ilustra en la figura 5.11 en la cual se observa al problema de explotación de información descrito a través de las notaciones definidas, desde una perspectiva orientada a las relaciones entre componentes.

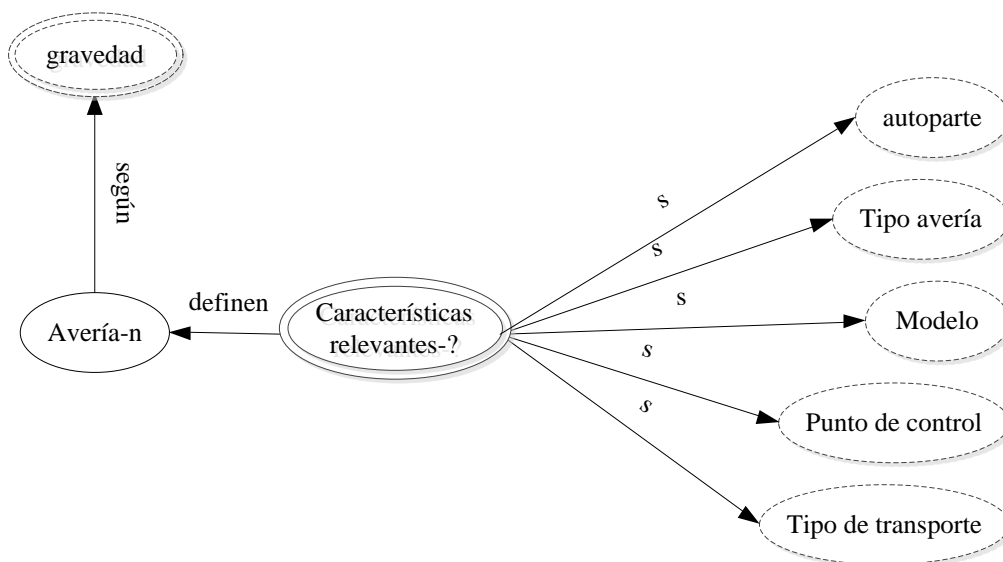


Figura 5.11. Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: industria automotriz

5.2.4. Aplicación de la Técnica Utilizada en la Fase de Determinación del Proceso de Explotación de Información

En esta sección se aplican al presente caso de validación la etapa “*derivación del proceso de explotación de información*” (sección 5.2.4.1), cuyo resultado esperado es el resultado final del proceso propuesto.

5.2.4.1. Paso 7: “Derivación del Proceso de Explotación de Información”

En este paso se implementan la técnica *algoritmo de derivación del proceso de explotación de información* (sección 5.2.4.1.1), cuyo objetivo es identificar el proceso de explotación de información a aplicar en base a las representaciones intermedias obtenidas. La Red Semántica del Modelo de Negocio (RSMN), generada en el paso 3, y la Red Semántica del Problema de Explotación de Información (RSPEI), generado en el paso 6, son los elementos de entrada del paso, siendo el elemento de salida el proceso de explotación de información a aplicar.

5.2.4.1.1. Técnica Algoritmo de Derivación del Proceso de Explotación de Información

Esta técnica está compuesta por un único paso, el cual está conformado por ocho subpasos condicionales, y genera a partir la Red Semántica del Modelo de Negocio (sección 5.2.2.3) y la Red Semántica del Problema de Explotación de Información (sección 5.2.3.3) aplicadas a la técnica propuesta, el proceso de explotación de información a aplicar.

Paso 1. Ejecución del Algoritmo de derivación: se identifica en las redes semánticas que conforman el conjunto de elementos de entrada de la técnica, los patrones definidos en cada subpaso del Algoritmo de Derivación del Proceso de Explotación de Información (ADPEI).

La secuencia de pasos obtenida de la ejecución del presente caso de estudio en el Algoritmo de Derivación del Proceso de Explotación de Información (ADPEI) se presenta a continuación, junto con el proceso de explotación de información identificado como resultado. Para la comprensión de los elementos detectados en cada subpaso, se aplica la siguiente notación: líneas punteadas para arcos o nodos identificados como respuesta en cada subpaso, y subrayado para texto.

Subpaso 7.1: Se identifica un nodo variable en la Red Semántica del Problema de Explotación de Información (figura 5.12).

✓ Resultado obtenido: Ir a Subpaso 7.2

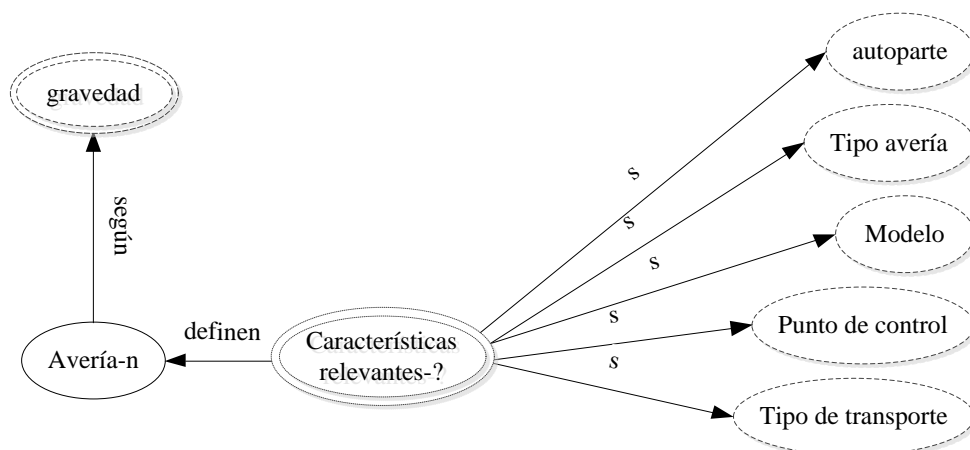


Figura 5.12. Identificación de subpaso 7.1 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: industria automotriz

Subpaso 7.2: No se identifica un nodo variable en la Red Semántica del Problema de Explotación de Información etiquetado con las palabras: pertenece, integra, es, es miembro de, u otras palabras de similar significado.

✓ Resultado obtenido: Ir a Subpaso 7.3

Subpaso 7.3: Se identifica un nodo variable, en la Red Semántica del Problema de Explotación de Información, etiquetado con las palabras “características relevantes” (figura 5.13).

✓ Resultado obtenido: **Aplicar Proceso Ponderación de Interdependencia de atributos**

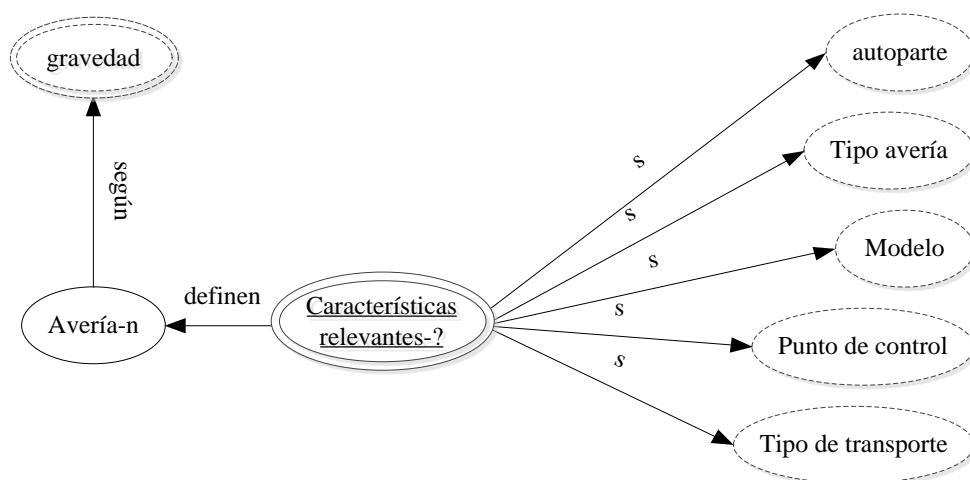


Figura 5.13. Identificación de subpaso 7.3 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: industria automotriz

Retornando como respuesta del paso y del *proceso de derivación de modelos* aplicado al caso de validación: *prestación de servicios de internet*, el texto:

“*Proceso de Descubrimiento de Reglas de Comportamiento*”.

5.3. CASO DE VALIDACIÓN: MARKET BASKET ANALYSIS

En esta sección se analiza el tercer caso de validación correspondiente a una empresa perteneciente a la industria automotriz. En la sección 5.3.1 se describe al modelo de negocio, sus datos y el problema de explotación de información a analizar. En la sección 5.3.2 se aplican las etapas y técnicas utilizadas para el desarrollo de las tareas correspondientes a la fase de *análisis del dominio del negocio*. En la sección 5.3.3 se aplican las etapas y técnicas utilizadas para el desarrollo de las tareas correspondientes a la fase de *análisis del problema de explotación de información*. En la sección 5.3.4 se aplican la etapa y técnica utilizada para el desarrollo de las tareas correspondientes a la fase de *determinación del proceso de explotación de información* obteniendo como resultado el proceso de explotación de información a aplicar en base al modelo del negocio y el problema analizado.

5.3.1. Descripción

En esta sección se presenta la información necesaria para comprender el modelo de negocio y el problema de explotación de información a analizar. El presente caso se obtiene de [Giudici, 2003], y se presentan las características relevantes del mismo:

“...*Descripción del Negocio y Problema de Negocio:*

Este caso de estudio observa el comportamiento de los clientes en los mercados, mediante el análisis de la canasta de compras. El objetivo del análisis de la canasta de compras es identificar aquellos productos que suelen venderse juntos en una misma compra. Este tipo de conocimiento puede ser muy importante, ya sea para reorganizar la distribución de los productos en el mercado, poniendo en las góndolas más cerca aquellos productos que se venden frecuentemente juntos. Además puede ser utilizado para mejorar la eficiencia de campañas de promoción: productos que están asociados no deberían ponerse en promoción durante el mismo periodo.

Se consideran todas las transacciones de una tienda representativa del sur de Italia entre el 2 de enero y el 21 de abril de 2001 (75 días). Cada transacción se identifica con un código único.

Entre los distintos productos que se venden a los clientes en la tienda, estos se agrupan en 493 categorías, de las cuales solo 10 serán las consideradas para este análisis, aquellas más vendidas.

El cliente posee un código de tarjeta que lo identifica del resto y permite asociar cada compra del cliente realizada, además de registrarse la fecha de la compra y los productos comprados.

Problema de explotación de información:

Determinar distintos grupos de productos relacionados en las compras e identificar aquellos que tengan un mayor nivel de incidencia sobre la compra de productos de dicho grupo.

Descripción de los datos:

Código de transacción: Identificador único de una compra. Tipo numérico.

Código de tarjeta: Identificador único de la tarjeta del cliente. Tipo numérico.

Fecha: fecha de la compra realizada. Formato mm/dd/aaaa

Producto: Categoría del producto. Las categorías a utilizar son: pasta, leche, agua, galletas, café, brioche, yogurt, vegetales congelados, atún y cerveza.

Cada producto tiene dos posibles valores:

- *No: no se compró dicho producto,*
- *Si: si el producto fue comprado al menos una vez...*

5.3.2. Aplicación de las Técnicas Utilizadas en la Fase de Análisis del Dominio del Negocio

En esta sección se aplican al caso de validación actual, las etapas y técnicas utilizadas para el desarrollo de las tareas correspondientes a la fase de análisis del dominio del negocio. Las etapas que componen la fase son: *identificación de los elementos y estructura del dominio* (sección 5.3.2.1), *identificación de relaciones entre conceptos del dominio* (sección 5.3.2.2) y *conceptualización del dominio* (sección 5.3.2.3).

5.3.2.1. Paso 1: “Identificación de los elementos y estructura del dominio”

En este paso se implementan dos técnicas: la *tabla término-categoría-definición del dominio* (sección 5.3.2.1.1), orientada a identificar los elementos relevantes del modelo de negocio y la *tabla concepto-atributo-relación-valor del dominio* (sección 5.3.2.1.2), cuyo objetivo es definir la estructura de los conceptos relevantes que integran al negocio. La descripción del dominio y los datos del negocio ingresan como elemento de entrada del paso, y las tablas término-categoría-definición del dominio (TCDD) y tabla concepto-atributo-relación-valor del dominio (CARVD), son los productos salida del paso.

5.3.2.1.1. Técnica Tabla Término-Categoría-Definición del Dominio

Esta técnica está compuesta por tres pasos, y genera como resultado de aplicar dichos pasos a la descripción del negocio provista en la sección 5.3.1, la tabla término-categoría-definición del dominio (TCDD).

Paso 1. *Identificación de términos relevantes:* se lleva a cabo el análisis de la descripción del negocio a los efectos de identificar y registrar alfabéticamente los elementos que permiten comprender el negocio.

"agua", "atún ", "brioche", "café", "cerveza", "Cliente", "código compra", "código tarjeta", "Compra", "de", "de", "de", "de", "de", "de", "de", "de", "de", "de", "de", "fecha", "galletas", "identifica", "identifica", "leche", "pasta", "realiza", "tiene", "vegetales congelados" y "yogurt".

A continuación se presenta el listado de términos identificados en el paso anterior, junto con la categoría asignada a cada uno de ellos, representando dicha asignación mediante pares ordenados, donde el primer elemento es el nombre del término y el segundo la categoría designada:

Paso 3. Describir el término: se lleva a cabo una descripción de cada término identificado, en base al dominio del negocio que integra.

Los datos producidos en los pasos precedentes se integran a la tabla Término-Categoría-Definición del Dominio (TCDD) junto con la descripción de cada concepto. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.13 en la cual se observa el conjunto de términos identificados, su categoría y la descripción que los define de acuerdo al negocio analizado.

Esta técnica está compuesta por cuatro pasos, cuyos elementos de entrada son la descripción de los datos del negocio provista en la sección 5.3.1, y la tabla término-categoría-definición del dominio (TCDD) obtenida de ejecutar la técnica tabla término-categoría-definición del dominio. El producto resultado del paso es la tabla concepto-atributo-relación-valor del dominio (CARVD).

Término	Categoría	Definición
agua	Atributo	producto que forma parte de la compra
atún	Atributo	producto que forma parte de la compra
brioche	Atributo	producto que forma parte de la compra
café	Atributo	producto que forma parte de la compra
cerveza	Atributo	producto que forma parte de la compra
Cliente	Concepto	Persona que compra al negocio
código compra	Atributo	identificador de cada compra
código tarjeta	Atributo	Identificador de cada cliente
Compra	Concepto	Venta de productos de la empresa
de	Relación	la compra puede ser de pasta
de	Relación	la compra puede ser de leche
de	Relación	la compra puede ser de agua
de	Relación	la compra puede ser de galletas
de	Relación	la compra puede ser de atún
de	Relación	la compra puede ser de café
de	Relación	la compra puede ser de brioche
de	Relación	la compra puede ser de yogurt
de	Relación	la compra puede ser de vegetales congelados
de	Relación	la compra puede ser de cerveza
fecha	Atributo	fecha de la compra
galletas	Atributo	producto que forma parte de la compra
identifica	Relación	el código de compra identifica a una compra
identifica	Relación	el código de tarjeta identifica a un cliente
leche	Atributo	producto que forma parte de la compra
pasta	Atributo	producto que forma parte de la compra
realiza	Relación	el cliente realiza una compra
tiene	Relación	la compra tiene una fecha
vegetales congelados	Atributo	producto que forma parte de la compra
yogurt	Atributo	producto que forma parte de la compra

Tabla 5.13. Tabla Término-Categoría-Definición del Dominio (TCDD) a partir del caso de validación: market basket analysis

Paso 1. Identificación de conceptos: a partir de los términos identificados en la tabla término-categoría-definición del dominio (TCDD), se registran aquellos elementos de categoría “Concepto” en la columna homónima de la tabla concepto-atributo-relación-valor del dominio (CARVD).

Se identifican dos elementos de categoría “Concepto”: Cliente y Compra.

Paso 2. Identificación de atributos por concepto: se lleva a cabo una integración de cada término cuya categoría es “Atributo” identificado en la tabla término-categoría-definición del dominio (TCDD), al concepto previamente identificado en la tabla concepto-atributo-relación-valor del dominio (CARVD).

A continuación se presenta una lista de pares ordenados, en donde el primer elemento corresponde al concepto identificado en el paso previo, y el segundo al atributo asociado al concepto:

(Cliente , código tarjeta), (Compra , código compra), (Compra , fecha), (Compra , pasta), (Compra , leche), (Compra , agua), (Compra , galletas), (Compra , café), (Compra , brioche), (Compra , yogurt), (Compra , vegetales congelados), (Compra , atún) y (Compra , cerveza).

Paso 3. *Identificación de la relación:* se identifica y registra aquel elemento de la tabla término-categoría-definición del dominio (TCDD) cuya categoría es “Relación” que relaciona al concepto con el atributo, registrándolo en la columna homónima del atributo correspondiente.

A continuación se presenta una lista de ternas ordenadas, cuyos primeros dos elementos son los identificados en el punto anterior, agregándose el nombre de la relación que los vincula:

(Cliente , código tarjeta , identifica), (Compra , código compra , identifica), (Compra , fecha , tiene), (Compra , pasta , de), (Compra , leche , de), (Compra , agua , de), (Compra , galletas , de), (Compra , café , de), (Compra , brioche , de), (Compra , yogurt , de), (Compra , vegetales congelados , de), (Compra , atún , de) y (Compra , cerveza , de).

Paso 4. *Identificación de los valores posibles:* a partir de la descripción de los datos, se definen los posibles valores de cada atributo identificado.

Los datos producidos en los pasos precedentes se integran a la tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) junto con los valores posibles de cada atributo. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.14 en la cual se observa al conjunto de términos identificados en la técnica anterior, presentados de una forma organizada la cual permite describir la estructura de cada concepto perteneciente al dominio del negocio.

5.3.2.2. Paso 2: “Identificación de Relaciones entre Conceptos del Dominio”

En este paso se implementan la técnica *tabla concepto-relación del dominio* (sección 5.3.2.2.1), orientada a identificar las interacciones entre los conceptos del dominio de negocio previamente identificados. Las tablas término-categoría-definición del dominio (TCDD) y concepto-atributo-relación-valor del dominio (CARVD), obtenidas a partir del paso 1, ingresan como elemento de entrada del paso actual, y la tabla concepto-relación del dominio (CRD) es su producto salida.

Concepto	Atributo	Relación	Valor
Cliente	código tarjeta	identifica	Numérico
Compra	código compra	identifica	Numérico
	fecha	tiene	mm/dd/aaaa
	pasta	de	Si/No
	leche	de	Si/No
	agua	de	Si/No
	galletas	de	Si/No
	café	de	Si/No
	brioche	de	Si/No
	yogurt	de	Si/No
	vegetales congelados	de	Si/No
	atún	de	Si/No
	cerveza	de	Si/No

Tabla 5.14. Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) a partir del caso de validación: market basket analysis

5.3.2.2.1. Técnica Tabla Concepto-Relación del Dominio

Esta técnica está compuesta por dos pasos, y genera como resultado de aplicar dichos pasos a las tablas término-categoría-definición del dominio (TCDD) y concepto-atributo-relación-valor del dominio (CARVD) obtenidas en la sección 5.3.2.1, la tabla concepto-relación del dominio (CRD).

Paso 1. Identificación de relaciones entre conceptos: se registran aquellos términos de categoría “Relación”, identificados en la tabla término-categoría-definición del dominio (TCDD), que relacionen a dos conceptos identificados, y se describe a dicha relación.

Se identifica al siguiente término como vinculador entre conceptos: “realiza”, cuya descripción es “el cliente realiza una compra”.

Paso 2. Identificación ordenada de conceptos relacionados: se registran los conceptos integrantes de las relaciones identificadas en el paso anterior, ingresando al concepto generador de la relación en la columna “Concepto” y al concepto asociado en la columna homónima.

El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.15 en la cual se observan los conceptos relacionados, su relación y una descripción de la misma en base al dominio del negocio.

Concepto	Concepto asociado	Relación	Descripción
Cliente	Compra	realiza	el cliente realiza una compra

Tabla 5.15. Tabla Concepto-Relación del Dominio (CRD) a partir del caso de validación: market basket analysis

5.3.2.3. Paso 3: “Conceptualización Del Dominio”

En este paso se implementan la técnica *red semántica del modelo de negocio* (sección 5.3.2.3.1), cuyo objetivo es conceptualizar la información del dominio del negocio, desde un enfoque orientado a las relaciones de sus componentes. Las tablas concepto-atributo-relación-valor del dominio (CARVD) y la tabla concepto-relación del dominio (CRD), obtenidas en los pasos 1 y 2 respectivamente, ingresan como elemento de entrada del paso actual, y la red semántica del modelo de negocio (RSMN) es su producto salida.

5.3.2.3.1. Técnica Red Semántica del Modelo de Negocio

Esta técnica está compuesta por dos pasos, y genera como resultado de aplicar dichos pasos a las tablas concepto-atributo-relación-valor del dominio (CARVD) y concepto-relación del dominio (CRD) obtenidas en la sección 5.3.2.1.2 y 5.3.2.2.1 respectivamente, la red semántica del modelo de negocio (RSMN).

Paso 1. Representación de los conceptos: se representa a partir de la notación definida en la sección 4.2, la estructura de los conceptos identificados en la tabla concepto-atributo-relación-valor del dominio (CARVD).

El producto obtenido de aplicar dicho paso al presente caso de estudio, se ilustra en la figura 5.14 en la cual se observa al modelo de negocio descrito a través de las notaciones definidas, desde una perspectiva orientada a las relaciones entre componentes.

Paso 2. Representación de las interacciones entre conceptos: se representan las relaciones entre conceptos identificadas en la tabla concepto-relación del dominio (CRD), mediante la anotación provista en la sección 4.2.

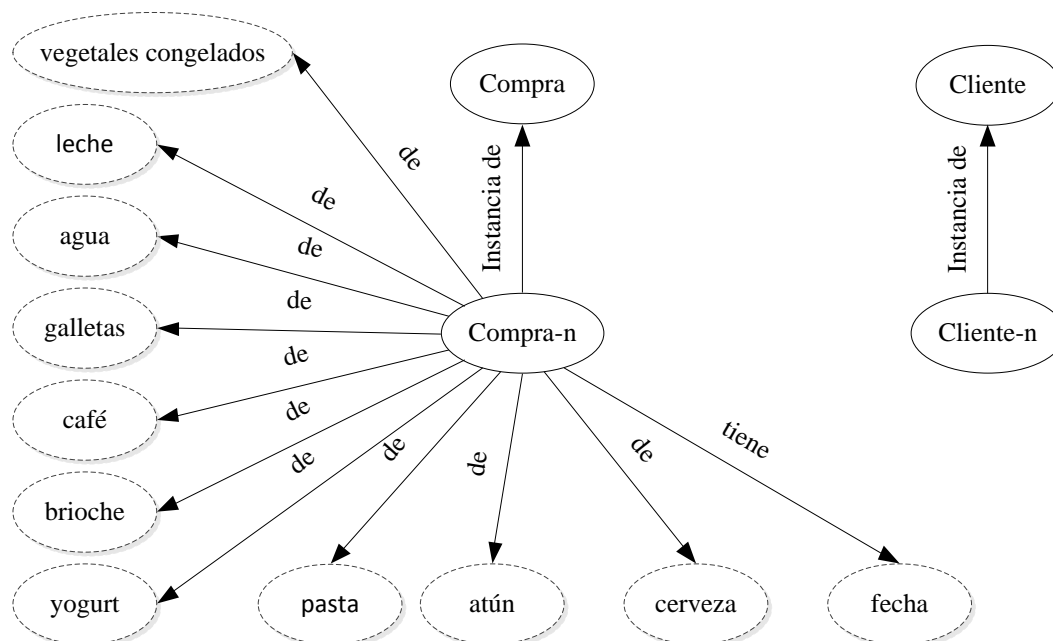


Figura 5.14. Paso 1: Red Semántica del Modelo de Negocio (RSMN) a partir del caso de validación: market basket analysis

El producto obtenido de integrar el segundo paso, al paso previo aplicando dichos pasos al presente caso de estudio, se ilustra en la figura 5.15 en la cual se observa al modelo del negocio descrito a través de las notaciones definidas, desde una perspectiva orientada a las relaciones entre componentes.

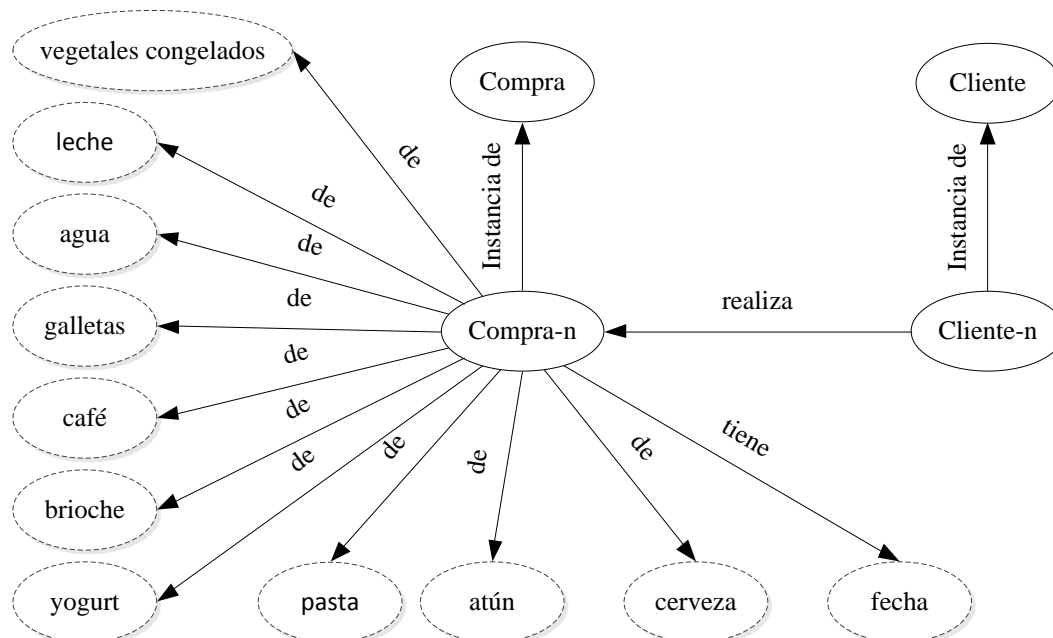


Figura 5.15. Red Semántica del Modelo de Negocio (RSMN) a partir del caso de validación: market basket analysis

5.3.3. Aplicación de las Técnicas Utilizadas en la Fase de Análisis del Problema de Explotación de Información

En esta sección se aplican al presente caso de validación las etapas y técnicas utilizadas para el desarrollo de las tareas correspondientes a la fase de análisis del problema de explotación de información. Las etapas que componen la fase son: *identificación de los elementos y estructura del problema de explotación de información* (sección 5.3.3.1), *identificación de relaciones entre conceptos del problema de explotación de información* (sección 5.3.3.2) y *conceptualización del problema de explotación de información* (sección 5.3.3.3).

5.2.3.1. Paso 4: “Identificación de los elementos y estructura del Problema de Explotación de Información”

En este paso se implementan dos técnicas: la *tabla término-categoría-definición del problema de explotación de información* (sección 5.3.3.1.1), orientada a identificar los elementos relevantes del problema de explotación de información y la *tabla concepto-atributo-relación-valor del problema de explotación de información* (sección 5.3.3.1.2), cuyo objetivo es definir la estructura de los conceptos relevantes que integran al problema. La descripción y los datos del problema ingresan como elemento de entrada del paso, y las tablas término-categoría-definición del problema de explotación de información (TCDPEI) y tabla concepto-atributo-relación-valor extendida del problema de explotación de información (CARVEPEI), son los productos salida del paso.

5.3.3.1.1. Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información

Esta técnica está compuesta por tres pasos, y genera como resultado de aplicar dichos pasos a la descripción del problema de explotación de información provista en la sección 5.3.1, la tabla término-categoría-definición del problema de explotación de información (TCDPEI).

Paso 1. *Identificación de términos relevantes:* se lleva a cabo el análisis de la descripción del problema de explotación de información a los efectos de identificar y registrar alfabéticamente los elementos que permiten comprender al mismo.

A continuación se presenta el listado de términos identificados, ordenado alfabéticamente:

Paso 2. Clasificación de términos por categoría: a partir de los términos identificados en el paso anterior, se los clasifica de acuerdo al tipo de elemento que representa (sección 4.3.2.1).

(agua , Atributo), (atún , Atributo), (brioche , Atributo), (café , Atributo), (cerveza , Atributo),
(código factores incidentes , Atributo), (código grupo , Atributo), (Compra , Concepto), (definen ,
Relación), (factores incidentes , Concepto), (galletas , Atributo), (Grupos , Concepto),
(identifica , Relación), (identifica , Relación), (integrado según , Relación), (leche , Atributo), (pasta
, Atributo), (según , Relación), (según , Relación), (según , Relación), (según , Relación), (según ,
Relación), (según , Relación), (según , Relación), (según , Relación), (según , Relación), (según ,
Relación), (Subconjunto , Relación), (Subconjunto , Relación), (Subconjunto , Relación),
(Subconjunto , Relación), (Subconjunto , Relación), (Subconjunto , Relación), (Subconjunto ,
Relación), (Subconjunto , Relación), (Subconjunto , Relación), (Subconjunto , Relación), (vegetales
congelados , Atributo) y (yogurt , Atributo).

Los datos producidos en los pasos precedentes se integran a la tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) junto con la descripción de cada concepto. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.16 en la cual se observa el conjunto de términos identificados, su categoría y la descripción que los define en base al problema de explotación de información analizado.

Término	Categoría	Definición
agua	Atributo	producto que forma parte de la compra
atún	Atributo	producto que forma parte de la compra
brioche	Atributo	producto que forma parte de la compra
café	Atributo	producto que forma parte de la compra
cerveza	Atributo	producto que forma parte de la compra
código factores incidentes	Atributo	identificador de los factores incidentes de cada grupo
código grupo	Atributo	identificador de cada grupo
Compra	Concepto	Venta de productos de la empresa
definen	Relación	los factores incidentes definen la compra de productos pertenecientes a un grupo determinado
factores incidentes	Concepto	conjunto de productos que inciden en la compra de determinados productos
galletas	Atributo	producto que forma parte de la compra
Grupos	Concepto	Conjunto de productos relacionados
identifica	Relación	el código de grupo identifica a un único grupo
identifica	Relación	el código de factores incidentes identifica el conjunto de factores que definen la compra de un conjunto de productos
integrado según	Relación	los grupos se integran según las compras
leche	Atributo	producto que forma parte de la compra
pasta	Atributo	producto que forma parte de la compra
según	Relación	la compra pertenece a un grupo según si compró pasta
según	Relación	la compra pertenece a un grupo según si compró leche
según	Relación	la compra pertenece a un grupo según si compró agua
según	Relación	la compra pertenece a un grupo según si compró galletas
según	Relación	la compra pertenece a un grupo según si compró café
según	Relación	la compra pertenece a un grupo según si compró brioche
según	Relación	la compra pertenece a un grupo según si compró yogurt
según	Relación	la compra pertenece a un grupo según si compró vegetales congelados
según	Relación	la compra pertenece a un grupo según si compró atún
según	Relación	la compra pertenece a un grupo según si compró cerveza
Subconjunto	Relación	la pasta puede ser un producto que defina la compra de elementos de un grupo
Subconjunto	Relación	la leche puede ser un producto que defina la compra de elementos de un grupo
Subconjunto	Relación	el agua puede ser un producto que defina la compra de elementos de un grupo
Subconjunto	Relación	las galletas puede ser un producto que defina la compra de elementos de un grupo
Subconjunto	Relación	el café puede ser un producto que defina la compra de elementos de un grupo
Subconjunto	Relación	la brioche puede ser un producto que defina la compra de elementos de un grupo
Subconjunto	Relación	el yogurt puede ser un producto que defina la compra de elementos de un grupo
Subconjunto	Relación	los vegetales congelados puede ser un producto que defina la compra de elementos de un grupo
Subconjunto	Relación	el atún puede ser un producto que defina la compra de elementos de un grupo
Subconjunto	Relación	la cerveza puede ser un producto que defina la compra de elementos de un grupo
vegetales congelados	Atributo	producto que forma parte de la compra
yogurt	Atributo	producto que forma parte de la compra

Tabla 5.16. Tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) a partir del caso de validación: market basket analysis

5.3.3.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información

Esta técnica está compuesta por cinco pasos, cuyos elementos de entrada son la descripción de los datos del problema de negocio y del problema de explotación de información, presentadas en la sección 5.3.1, y la tabla término-categoría-definición del problema de explotación de información (TCDPEI) obtenida al ejecutar la técnica tabla término-categoría-definición del problema de explotación de información. El producto resultado del paso es la tabla concepto-atributo-relación-valor extendida del problema de explotación de información (CARVEPEI).

Paso 1. *Identificación de conceptos:* a partir de los términos identificados en la tabla término-categoría-definición del problema de explotación de información (TCDPEI), se registran aquellos elementos de categoría “Concepto” en la columna homónima de la tabla concepto-atributo-relación-valor extendida del problema de explotación de información (CARVEPEI).

Se identifican tres conceptos: “Compra”, “Grupo” y “Factores incidentes”.

Paso 2. *Identificación de atributos por concepto:* se lleva a cabo una integración de cada término cuya categoría es “Atributo” identificado en la tabla término-categoría-definición del problema de explotación de información (TCDPEI), al concepto previamente identificado en la tabla concepto-atributo-relación-valor extendida del problema de explotación de información (CARVEPEI).

A continuación se presenta una lista de pares ordenados, en donde el primer elemento corresponde al concepto identificado en el paso previo, y el segundo al atributo asociado a dicho concepto:

(Compra , pasta), (Compra , leche), (Compra , agua), (Compra , galletas), (Compra , café), (Compra , brioche), (Compra , yogurt), (Compra , vegetales congelados), (Compra , atún), (Compra , cerveza), (Grupo , código grupo), (Factores incidentes , código factores incidentes), (Factores incidentes , pasta), (Factores incidentes , leche), (Factores incidentes , agua), (Factores incidentes , galletas), (Factores incidentes , café), (Factores incidentes , brioche), (Factores incidentes , yogurt), (Factores incidentes , vegetales congelados), (Factores incidentes , atún) y (Factores incidentes , cerveza).

Paso 3. *Identificación de la relación:* se identifica y registra aquel elemento de la término-categoría-definición del problema de explotación de información (TCDPEI) cuya categoría es “Relación” que relaciona al concepto con el atributo, registrándolo en la columna homónima del atributo correspondiente.

A continuación se presenta una lista de ternas ordenadas, cuyos primeros dos elementos son los identificados en el punto anterior, agregándose el nombre de la relación que los vincula:

(Compra , pasta , según), (Compra , leche , según), (Compra , agua , según), (Compra , galletas , según), (Compra , café , según), (Compra , brioche , según), (Compra , yogurt , según), (Compra , vegetales congelados , según), (Compra , atún , según), (Compra , cerveza , según), (Grupo , código grupo , identifica), (Factores incidentes , código factores incidentes , identifica), (Factores incidentes , pasta , subconjunto), (Factores incidentes , leche , subconjunto), (Factores incidentes , agua , subconjunto), (Factores incidentes , galletas , subconjunto), (Factores incidentes , café , subconjunto), (Factores incidentes , brioche , subconjunto), (Factores incidentes , yogurt , subconjunto), (Factores incidentes , vegetales congelados , subconjunto), (Factores incidentes , atún , subconjunto) y (Factores incidentes , cerveza , subconjunto).

Paso 4. *Identificación de los valores posibles:* a partir de la descripción de los datos, se definen los posibles valores de cada atributo identificado.

Se presentan una lista de cuaternas cuyos primeros tres elementos son los identificados en el paso previo, agregando como cuarto elemento los valores posibles de cada atributo:

(Compra , pasta , según , Si/No), (Compra , leche , según , Si/No), (Compra , agua , según , Si/No), (Compra , galletas , según , Si/No), (Compra , café , según , Si/No), (Compra , brioche , según , Si/No), (Compra , yogurt , según , Si/No), (Compra , vegetales congelados , según , Si/No), (Compra , atún , según , Si/No), (Compra , cerveza , según , Si/No), (Grupo , código grupo , identifica , Numérico), (Factores incidentes , código factores incidentes , identifica , Numérico), (Factores incidentes , pasta , subconjunto , Si/No), (Factores incidentes , leche , subconjunto , Si/No), (Factores incidentes , agua , subconjunto , Si/No), (Factores incidentes , galletas , subconjunto , Si/No), (Factores incidentes , café , subconjunto , Si/No), (Factores incidentes , brioche , subconjunto , Si/No), (Factores incidentes , yogurt , subconjunto , Si/No), (Factores incidentes , vegetales congelados , subconjunto , Si/No), (Factores incidentes , atún , subconjunto , Si/No) y (Factores incidentes , cerveza , subconjunto , Si/No).

Paso 5. Identificación de los atributos de entrada y salida: a partir de la descripción del problema de explotación de información, se definen cuales atributos son valores de entrada y cuales son valores esperados o salida del problema.

Los datos producidos en los pasos precedentes se integran a la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) junto con los elementos identificados como entrada ó salida del problema. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.17 en la cual se observa al conjunto de términos identificados en la técnica anterior, presentados de una forma organizada la cual permite describir la estructura de cada concepto perteneciente al problema de explotación de información.

Concepto	Atributo	Relación	Entrada/Salida	Valor
Compra	pasta	según	Entrada	Si/No
	leche	según	Entrada	Si/No
	agua	según	Entrada	Si/No
	galletas	según	Entrada	Si/No
	café	según	Entrada	Si/No
	brioche	según	Entrada	Si/No
	yogurt	según	Entrada	Si/No
	vegetales congelados	según	Entrada	Si/No
	atún	según	Entrada	Si/No
	cerveza	según	Entrada	Si/No
Grupo	código grupo	identifica	Salida	Numérico
Factores incidentes	código factores incidentes	identifica	Salida	Numérico
	pasta	subconjunto		Si/No
	leche	subconjunto		Si/No
	agua	subconjunto		Si/No
	galletas	subconjunto		Si/No
	café	subconjunto		Si/No
	brioche	subconjunto		Si/No
	yogurt	subconjunto		Si/No
	vegetales congelados	subconjunto		Si/No
	atún	subconjunto		Si/No
	cerveza	subconjunto		Si/No

Tabla 5.17. Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) a partir del caso de validación: market basket analysis

5.3.3.2. Paso 5: “Identificación de relaciones entre conceptos del Problema de Explotación de Información”

En este paso se implementan la técnica *tabla concepto-relación del problema de explotación de información* (sección 5.3.3.2.1), orientada a identificar las interacciones entre los conceptos del problema de explotación de información previamente identificados. Las tablas término-categoría-

definición del problema de explotación de información (TCDPEI) y Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI), obtenidas a partir del paso 4, ingresan como elemento de entrada del paso actual, y la tabla concepto-relación del problema de explotación de información (CRPEI) es su producto salida.

5.3.3.2.1. Técnica Tabla Concepto-Relación del Problema de Explotación de Información

Esta técnica está compuesta por dos pasos, y genera como resultado de aplicar dichos pasos a las tablas término-categoría-definición del problema de explotación de información (TCDPEI) y Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) obtenidas en la sección 5.3.3.1, la tabla Concepto-Relación del Problema de Explotación de Información (CRPEI).

Paso 1. *Identificación de relaciones entre conceptos:* se registran aquellos términos de categoría “Relación”, identificados en la tabla término-categoría-definición del problema de explotación de información (TCDPEI), que relacionen a dos conceptos identificados, y se describe a dicha relación.

Se identifican dos términos de categoría “Relación” que vinculan conceptos: “definen” e “integrado según”, cuyas descripciones son “los factores incidentes definen la compra de productos pertenecientes a un grupo determinado” y “los grupos se integran según las compras” respectivamente.

Paso 2. *Identificación ordenada de conceptos relacionados:* se registran los conceptos integrantes de las relaciones identificadas en el paso anterior, ingresando al concepto generador de la relación en la columna “Concepto” y al concepto asociado en la columna homónima.

Los datos producidos en el paso previo se integran a la tabla Concepto-Relación del Problema de Explotación de Información (CRPEI) junto con los valores conceptos vinculados. El producto obtenido de aplicar dichos pasos al presente caso de estudio, se ilustra en la tabla 5.18 en la cual se observan las relaciones entre los distintos conceptos que componen al problema.

Concepto	Concepto asociado	Relación	Descripción
factores incidentes	Grupo	definen	los factores incidentes definen la compra de productos pertenecientes a un grupo determinado
Grupos	Compra	integrado según	los grupos se integran según las compras

Tabla 5.18. Tabla Concepto-Relación del Problema de Explotación de Información (CRPEI) a partir del caso de validación: market basket analysis

5.3.3.3. Paso 6: “Conceptualización del Problema de Explotación de Información”

En este paso se implementan la técnica *Red Semántica del Problema de Explotación de Información* (sección 5.3.3.3.1), cuyo objetivo es conceptualizar la información del problema de explotación de información, desde un enfoque orientado a las relaciones de sus componentes. Las tablas Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) y la Concepto-Relación del Problema de Explotación de Información (CRPEI), obtenidas en los pasos 4 y 5, ingresan como elemento de entrada del paso actual, y la Red Semántica del Problema de Explotación de Información (RSPEI) es su producto salida.

5.3.3.3.1. Técnica Red Semántica del Problema de Explotación de Información

Esta técnica está compuesta por tres pasos, y genera como resultado de aplicar dichos pasos a las tablas Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) y la Concepto-Relación del Problema de Explotación de Información (CRPEI) obtenidas en la sección 5.3.3.1.2 y 5.3.3.2.1 respectivamente, la Red Semántica del Problema de Explotación de Información (RSPEI).

Paso 1. Representación de los conceptos: se representa a partir de la notación definida en la sección 4.2, la estructura de los conceptos identificados en la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI).

Como resultado de aplicar el primer paso se obtiene la figura 5.16.

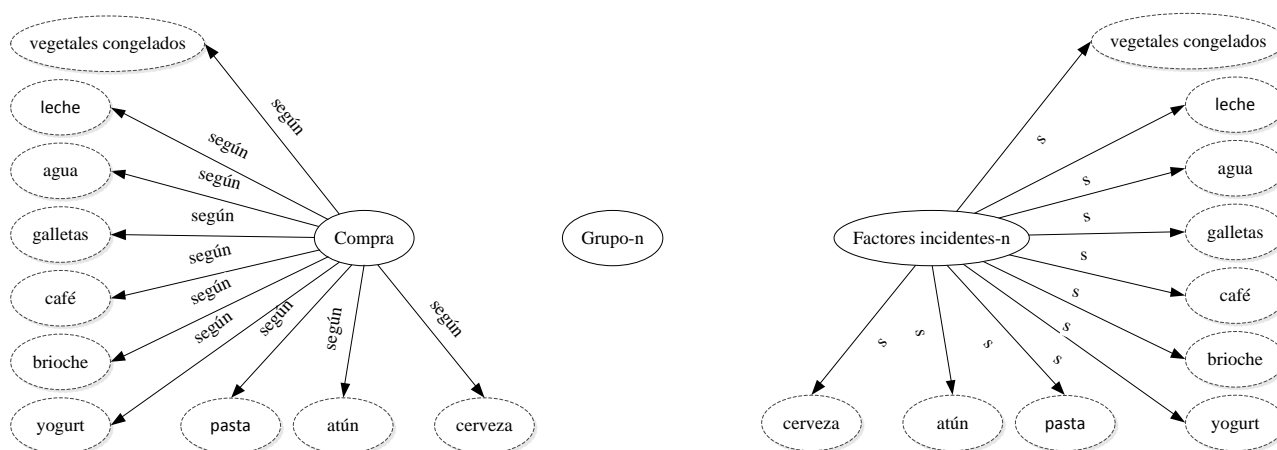


Figura 5.16. Paso 1: Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: market basket analysis

Paso 2. Representación de las interacciones entre conceptos: se representan las relaciones entre conceptos identificadas en la tabla Concepto-Relación del Problema de Explotación de Información (CRPEI), mediante la anotación provista en la sección 4.2.

Como resultado de integrar el segundo paso al paso previo se obtiene la figura 5.17.

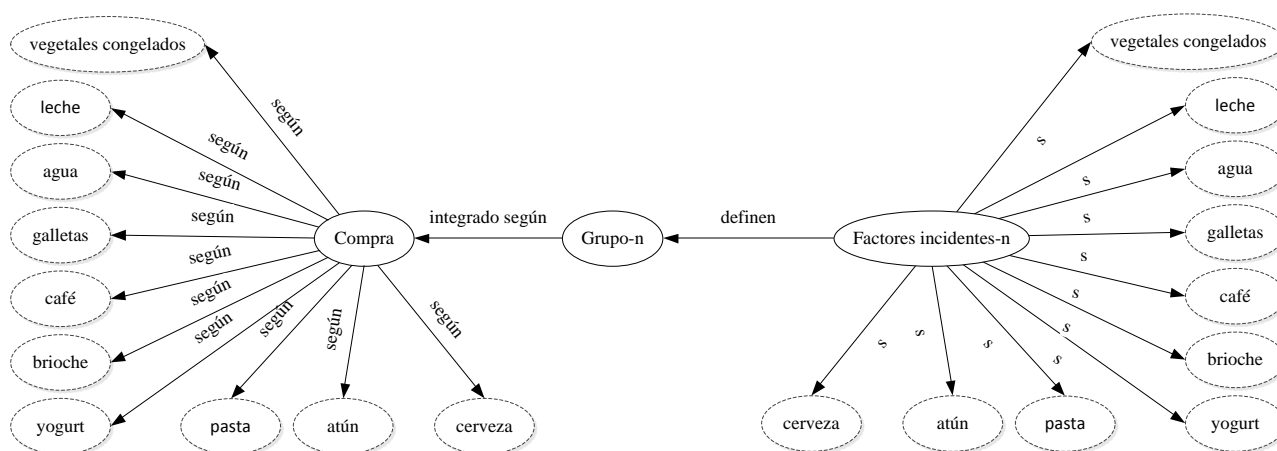


Figura 5.17. Paso 2: Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: market basket analysis

Paso 3. Representación de los elementos de entrada y salida: se representan los elementos de entrada y salida identificados en la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI), mediante la anotación propuesta en la sección 4.2.2.

El producto obtenido de integrar el tercer paso a los pasos previos, aplicando dichos pasos al presente caso de estudio, se ilustra en la figura 5.18 en la cual se observa al problema de

explotación de información descrito a través de las notaciones definidas, desde una perspectiva orientada a las relaciones entre componentes.

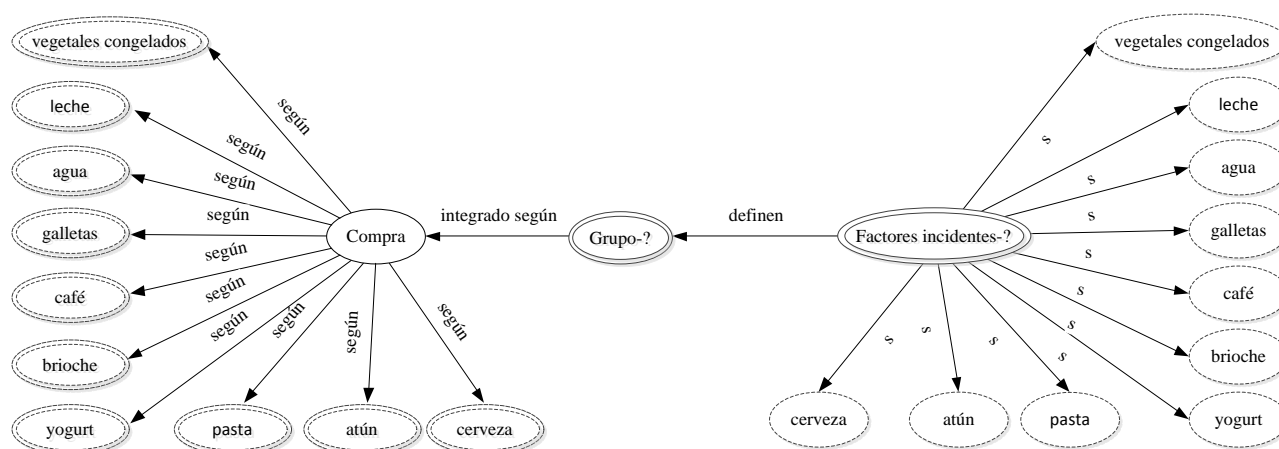


Figura 5.18. Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: market basket analysis

5.3.4. Aplicación de la Técnica Utilizada en la Fase de Determinación del Proceso de Explotación de Información

En esta sección se aplican al presente caso de validación la etapa “*derivación del proceso de explotación de información*” (sección 5.3.4.1), cuyo resultado esperado es el resultado final del proceso propuesto.

5.3.4.1. Paso 7: “Derivación del Proceso de Explotación de Información”

En este paso se implementan la técnica *algoritmo de derivación del proceso de explotación de información* (sección 5.3.4.1.1), cuyo objetivo es identificar el proceso de explotación de información a aplicar en base a las representaciones intermedias obtenidas. La Red Semántica del Modelo de Negocio (RSMN), generada en el paso 3, y la Red Semántica del Problema de Explotación de Información (RSPEI), generado en el paso 6, son los elementos de entrada del paso, siendo el elemento de salida el proceso de explotación de información a aplicar.

5.3.4.1.1. Técnica Algoritmo de Derivación del Proceso de Explotación de Información

Esta técnica está compuesta por un único paso, el cual está conformado por ocho subpasos condicionales, y genera a partir la Red Semántica del Modelo de Negocio (sección 5.3.2.3) y la Red

Semántica del Problema de Explotación de Información (sección 5.3.3.3) aplicadas a la técnica propuesta, el proceso de explotación de información a aplicar.

Paso 1. Ejecución del Algoritmo de derivación: se identifica en las redes semánticas que conforman el conjunto de elementos de entrada de la técnica, los patrones definidos en cada subpaso del Algoritmo de Derivación del Proceso de Explotación de Información (ADPEI).

La secuencia de pasos obtenida de la ejecución del presente caso de estudio en el Algoritmo de Derivación del Proceso de Explotación de Información (ADPEI) se presenta a continuación, junto con el proceso de explotación de información identificado como resultado. Para la comprensión de los elementos detectados en cada subpaso, se aplica la siguiente notación: líneas punteadas para arcos o nodos identificados como respuesta en cada subpaso, y subrayado para texto.

Subpaso 7.1: Se identifican dos nodos variables en la Red Semántica del Problema de Explotación de Información (figura 5.19).

✓ Resultado obtenido: Ir a Subpaso 7.5

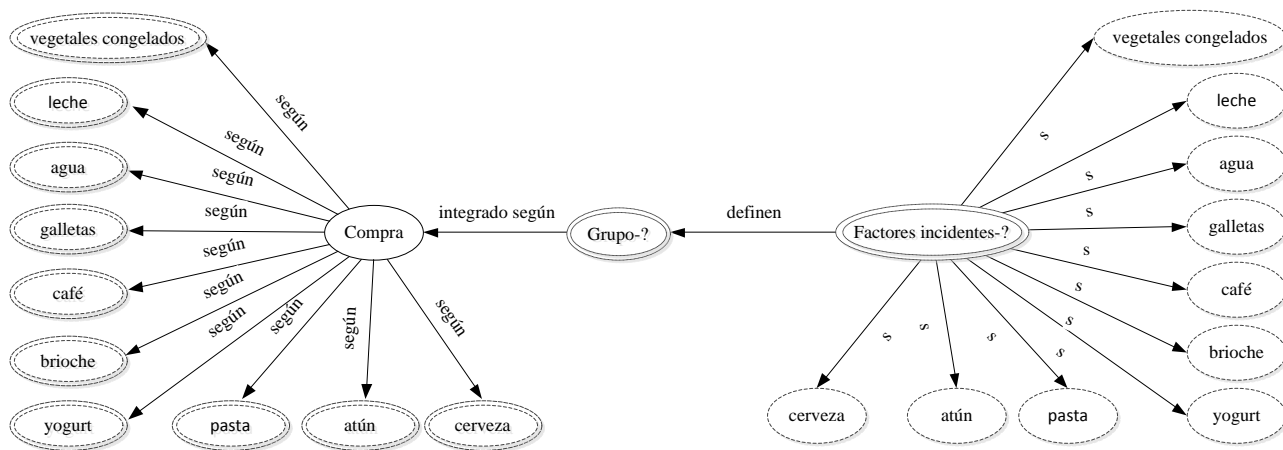


Figura 5.19. Identificación de subpaso 7.1 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: market basket analysis

Subpaso 7.5: Se identifica una relación etiquetada con las palabras “integrado según”, el cual tiene como nodo destino, uno de los nodos variables identificados en el paso 7.1, el cual se identifica con el otro nodo variable identificado (figura 5.20).

✓ Resultado obtenido: Ir a Subpaso 7.6

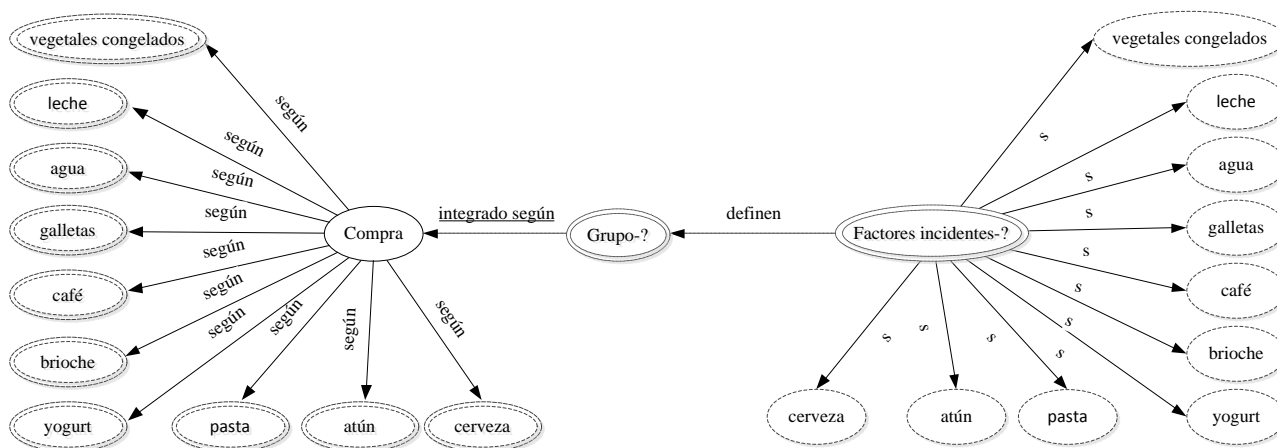


Figura 5.20. Identificación de subpaso 7.5 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: market basket analysis

Subpaso 7.6: El nodo variable a partir de la cual surge la relación entre nodos variables identificada en el paso anterior, se etiqueta con las palabras “factores incidentes” (figura 5.21).

✓ Resultado obtenido: Ir a Subpaso 7.7

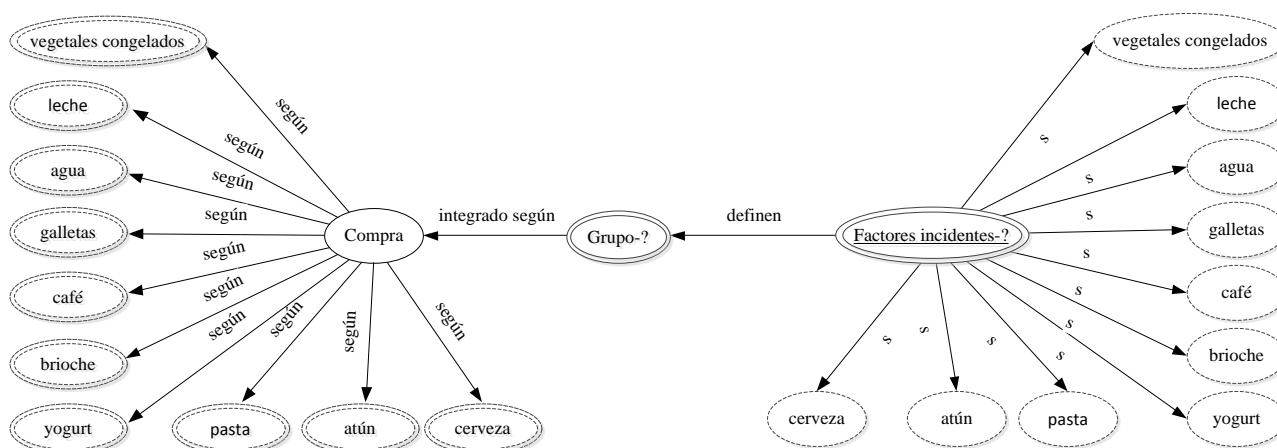


Figura 5.21. Identificación de subpaso 7.6 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de validación: market basket analysis

Subpaso 7.7: El nodo variable destino que integra la relación identificada en el subpaso 7.5 no es un nodo existente en la Red Semántica del Modelo del Negocio (RSMN).

✓ Resultado obtenido: **Aplicar Proceso Ponderación de Reglas de Pertenencia a Grupos (Grupos no definidos).**

Retornando como respuesta del paso y del *proceso de derivación de modelos* aplicado al caso de validación: market basket analysis, el texto:

“Aplicar Proceso Ponderación de Reglas de Pertenencia a Grupos (Grupos no definidos).”.

6. CONCLUSIONES

En este Capítulo se presentan las aportaciones de este trabajo (sección 6.1) y se destacan las futuras líneas de investigación que se consideran de interés en base al problema abierto presentado (sección 6.2).

6.1. APORTES DEL TRABAJO FINAL DE LICENCIATURA

En este Trabajo Final de Licenciatura se ha definido un proceso que permite partiendo de la descripción del dominio del negocio, del problema de explotación de información, y la descripción de sus datos, derivar el proceso de explotación de información a aplicar que responda al problema de negocio planteado por el cliente, permitiendo realizar dicho proceso de manera controlada y documentada. Además se han propuesto una serie formalismos de representación de conocimiento que permiten documentar la información necesaria para llevar a cabo la ejecución del *Proceso de Derivación de Modelos* propuesto, a partir del cual se identifica el proceso de explotación de información a aplicar.

En este contexto, en este TFL se ha propuesto:

- I. Un modelo de proceso de derivación del proceso de explotación de información que se desarrolla en tres fases: orientada al análisis del dominio del negocio, orientada al análisis del problema de explotación de información y orientada a determinar el proceso de explotación de información.
- II. Para la Fase Orientada al Análisis del Dominio del Negocio, se han propuesto las siguientes etapas: [i] *Identificación de los elementos y estructura del dominio*, la cual necesita la descripción del dominio del negocio y los datos del negocio como producto de entrada y proporciona como producto de salida la tabla Término-Categoría-Definición del Dominio (TCDD) y la tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD); [ii] *Identificación de relaciones entre conceptos del dominio*, que toma como producto de entrada a la tabla Término-Categoría-Definición del Dominio (TCDD) y la tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) y proporciona como producto de salida la tabla Concepto-Relación del Dominio (CRD); y [iii] *Conceptualización del dominio*, cuyos elementos de entrada son la tabla Concepto-Atributo-Relación-Valor del Dominio

- (CARVD) y la tabla Concepto-Relación del Dominio (CRD), y proporciona como producto de salida la Red Semántica del Modelo de Negocio (RSMN).
- III. Para la Fase Orientada al Análisis del Problema de Explotación de Información, se han propuesto las siguientes etapas: [iv] *Identificación de los elementos y estructura del Problema de Explotación de Información*, la cual necesita como productos de entrada la descripción del problema de explotación de información y los datos del problema de negocio y produce como producto de salida la tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) y la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI); [v] *Identificación de relaciones entre conceptos del Problema de Explotación de Información*, que tiene como elementos de entrada la tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) y la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI), y produce como salida la tabla Concepto-Relación del Problema de Explotación de Información (CRPEI); [vi] *Conceptualización del Problema de Explotación de Información* que tiene como elementos de entrada la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) y la tabla Concepto-Relación del Problema de Explotación de Información (CRPEI), y cuyo elemento de salida es la Red Semántica del Problema de Explotación de Información (RSPEI).
- IV. Para la Fase de Determinación del Proceso de Explotación de Información, se ha propuesto la siguiente etapa: [vii] *Derivación del Proceso de Explotación de Información*, la cual tiene como elementos de entrada la Red Semántica del Modelo de Negocio (RSMN) y la Red Semántica del Problema de Explotación de Información (RSPEI), y produce como elemento de salida el proceso de explotación de información a aplicar.
- V. Para la Fase Orientada al Análisis del Dominio del Negocio, se han desarrollado las siguientes técnicas: Técnica Tabla Término-Categoría-Definición del Dominio, Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio, Técnica Tabla Concepto-Relación del Dominio y Técnica Red Semántica del Modelo de Negocio.
- VI. Para la Fase Orientada al Análisis del Problema de Explotación de Información, se han desarrollado las siguientes técnicas: Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información, Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información, Técnica Tabla Concepto-Relación del Problema de Explotación de Información y Técnica Red Semántica del Problema de Explotación de Información.

VII. Para la Fase de Determinación del Proceso de Explotación de Información, se ha desarrollado la Técnica Algoritmo de Derivación del Proceso de Explotación de Información.

La propuesta del *proceso de derivación de modelos*, las etapas y técnicas asociadas han sido validadas en tres dominios de conocimiento con características bien diferenciadas: el primero sobre una empresa de prestación de servicios de internet, el segundo correspondiente a un negocio perteneciente a la industria automotriz y el tercero correspondiente a un análisis de la canasta de compras (“market basket analysis”).

6.2. FUTURAS LÍNEAS DE INVESTIGACIÓN

Durante el desarrollo de este proyecto de TFL han surgido cuestiones que si bien no son centrales al tema abordado en la misma, constituyen temas concomitantes que (en opinión del autor) darían lugar a las siguientes futuras líneas de investigación:

- Si bien el proceso propuesto en el trabajo aporta sistematicidad a la derivación del proceso de explotación de información y el mismo ha sido validado en dominios representativos, quedan como líneas de trabajos abiertas:
 - ♦ La validación empírica más amplia del proceso propuesto mediante la técnica de muestras apareadas basadas en grupos experimental y de control.
 - ♦ La validación empírica del proceso y de las técnicas propuestas en un conjunto vasto y representativo, considerando las distintas complejidades de dominios y problemas del negocio.
- Si bien el algoritmo de derivación propuesto prevé la existencia de errores como una posibilidad en las representaciones realizadas, queda como línea de trabajo abierta la identificación de un proceso que permita a partir de la obtención de “*Representación errónea*” como resultado de aplicar la Técnica Algoritmo de Derivación del Proceso de Explotación de Información, detectar el o los errores en los formalismos de representación realizados.

7. REFERENCIAS

- Basso, D. y Vegega, C. 2013. *Caso de Estudio: Telefonía Celular. Trabajo Final del Seminario de Inteligencia de Negocio*. Maestría en Ingeniería de Sistemas de Información. Escuela de Posgrado. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina.
- Britos, P. y García Martínez, R. 2009. *Propuesta de Procesos de Explotación de Información*. Proceedings XV Congreso Argentino de Ciencias de la Computación Workshop de Base de Datos y Minería de Datos. Págs. 1041-1050. ISBN 978-897-24068-4-1
- Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 1999. *CRISP-DM 1.0 Step by step data mining guide*.
- Dasgupta, S., Vankayala, K. 2007. *Developing Realtime Business Intelligence Systems. The Agile Way*. Proceedings 1st Annual 2007 IEEE Systems Conference. Pág. 63-69.
- Fayad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uhturudsamy, R. 1996. *Advances in Knowledge Discovery and Data Mining, (editors)*. AAAI Press.
- Flores, H. D. 2009. *Detección de Patrones de Daños y Averías en la Industria Automotriz*. Tesis de Magister en Ingeniería en Sistemas de Información.
- García Martínez, R. y Britos, P. 2004. *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería. ISBN 987-1104-15-4.
- García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell, J. 2011. *Towards an Information Mining Engineering. En Software Engineering, Methods, Modeling and Teaching*. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6. Páginas 83-99.
- García-Martínez, R., Britos, P., Rodríguez, D. 2013. *Information Mining Processes Based on Intelligent Systems*. Lecture Notes on Artificial Intelligence, 7906: 402-410. ISBN 978-3- 642-38576-6.
- García Martínez, R., Servente, M. y Pasquini, D. 2003. *Sistemas Inteligentes*. Editorial Nueva Librería. Buenos Aires. ISBN 987-1104-05-7.

- Giudici, Paolo. 2003. *Applied Data Mining: statistical methods for business and industry*. ISBN 0-470-84679-8.
- Grigori, D., Casati, F., Castellanos, M., Dayal, u., Sayal, M., Shan, M. 2004. *Business Process Intelligence*. *Computers in Industry* 53(3): 321-343.
- Grigori, D., Casati, F., Dayal, U., Shan, M. 2001. *Improving Business Process Quality through Exception Understanding, Prediction, and Prevention*. *Proceedings 27th International Conference on Very Large Data Bases*. Pág. 159-168.
- Grossman, R., Kasif, S., Moore, R., Rocke, D., Ullman, J. 1999. *Data Mining Research: Opportunities and Challenges*.
- Koubarakis, M., Plexousakis, D. 2000. *A Formal Model for Business Process Modeling and Design*. *Lecture Notes in Computer Science*, 1789: 142-156.
- Langseth, J., Vivatrat, N. 2003. *Why Proactive Business Intelligence is a Hallmark of the Real-Time Enterprise: Outward Bound*. *Intelligent Enterprise* 5(18): 34-41.
- Larose, D. T. 2005. *Discovering Knowledge in Data, an introduction to DataMining*. John Wiley & Sons, inc. Estados Unidos.
- Lönnqvist, A., Pirttimäki, V. 2006. *The Measurement of Business Intelligence*. *Information Systems Management*, 23(1): 32-40.
- Mäkipää, M. 2004. *The Role and Types of Business Information in Different "Schools of Thought" of Strategic Management*. *Proceedings Frontiers of E-Business Research*. Pág. 414-427.
- Morik, K., Rüping, S. 2002. *A Multistrategy Approach to the Classification of Phases in Business Cycles*. *Lecture Notes in Computer Science*, 2430: 307-318.
- Moss, L. 2003. *Nontechnical Infrastructure of BI Applications*. *DM Review* 13(1): 42-45.
- Negash, S., Gray, P. 2008. *Business Intelligence*. En *Handbook on Decision Support Systems 2*, ed. F. Burstein y C. Holsapple (Heidelberg, Springer), Pág. 175-193.
- Nguyen, T., Schiefer, J., Tjoa, M. 2005. *Sense & Response Service Architecture (SARESA): An Approach Towards a Real-Time Business Intelligence Solution and Its Use for a*

Fraud Detection Application. Proceedings of the 8th ACM International Workshop on Data Warehousing and OLAP. Pág. 77-86.

Osterwalder, A. 2004. *The Business Model Ontology a Proposition in a Design Science Approach*. Tesis Doctoral. Universidad de Lausanne.

Pirttimäki, V., Lönnqvist, A., Karjaluo, A. 2006. *Measurement of Business Intelligence in a Finnish Telecommunications Company*. Electronic Journal of Knowledge Management, 4(1): 83-90.

Reinschmidt, J., Allison F. 2000. *Business Intelligence Certification Guide*. IBM International Technical Support Organization. SG24-5747-00.

Reiterer, H., Mußler, G., Mann, T., Handschuh, S. 2000. *INSYDER: An Information Assistant for Business Intelligence*. Proceedings 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Pág.112-119.

Rodríguez, D. 2012. *Espacios virtuales para la formación de investigadores elementos de análisis y diseño*. Tesis de Maestría. Universidad Nacional de La Plata.

Rodríguez, D., García Martínez, R. 2012. *Modeling the Interactions in Virtual Spaces Oriented to Collaborative Work*. Capítulo 10 en Software Engineering: Methods, Modeling, and Teaching, Volume 2. Pág. 79-84. Sello Editorial de la Pontificia Universidad Católica del Perú. ISBN 978-612-4057-84-7.

Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2010. *Estimación Empírica de Carga de Trabajo en Proyectos de Explotación de Información*. Anales del XVI Congreso Argentino de Ciencias de la Computación. Pág. 664-673. ISBN 978-950-9474-49-9.

Rudin, K., Cressy, D. 2003. *Will the Real Analytic Application Please Stand Up?* DM Review 13(3): 30-34.

Sánchez, J. P. 2012. *Propuesta de Estandarización de Procesos*. Tesis de Maestro en Administración.

Schiefer, J., Jeng, J., Kapoor, S., Chowdhary, P. 2004. *Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence*. Proceedings 2004 IEEE International Conference on E-Commerce Technology. Pág. 162-169.

- Sommerville, I. 2005. *Ingeniería del software séptima edición*.
- Sowa, J. 1992. *Semantic Networks*. En Encyclopedia of Artificial intelligence (Editor S. Shapiro, 2da edición). Wiley & Sons. ISBN 978-0471503071.
- Stefanovic, N., Majstorovic. V., Stefanovic, D. 2006. *Supply Chain Business Intelligence Model*. Proceedings 13th International Conference on Life Cycle Engineering. Pág. 613-618.
- Suárez-Figueroa, M. y Gómez-Pérez, A. 2008. *Redes Semánticas y Marcos*. En: *Inteligencia Artificial: Técnicas, Métodos y Aplicaciones*. McGraw Hill. Pág. 131-170. ISBN 978-84-481-5618-3.
- Thomas, J. 2001. *Business Intelligence - Why?*. eAI Journal, Julio: 47-49.
- Thomsen, E. 2003. *BI's Promised Land*. Intelligent Enterprise, 6(4): 21-25.
- Vanrell, J. A. 2011. *Un Modelo de Procesos para Proyectos de Explotación de Información*. Tesis de Maestría en Ingeniería en Sistemas de Información. Escuela de Postgrado FRBA-UTN.
- Winston, P. 1994. *Inteligencia Artificial (3ra edición)*. Addison Wesley Iberoamericana. ISBN 0-201-51876-7.